G. Bamberg · K. Spremann (Eds.)

Agency Theory, Information, and Incentives

Springer-Verlag

Günter Bamberg and Klaus Spremann (Eds.)

Agency Theory, Information, and Incentives

With Contributions by
W. Ballwieser, G. Bamberg, M. J. Beckmann, H. Bester
M. Blickle, R. Ewert, G. Feichtinger, V. Firchau, F. Fricke
H. Funke, M. Gaynor, A. Gruber, T. Hartmann-Wendels
M. Hellwig, T. Kaulmann, P. R. Kleindorfer, J. P. Krahnen
G. Meran, H. Milde, A. Pfingsten, E. Schanze
R. H. Schmidt, D. Schneider, G. Sorger, K. Spremann
P. Swoboda, J. Voeller, A. Wagenhofer, J. E. M. Wilhelm

Corrected Reprint of the First Edition

With 37 Figures and 7 Tables

Springer-Verlag Berlin Heidelberg New York
London Paris Tokyo Hong Kong

Prof. Dr. Günter Bamberg
Lehrstuhl für Statistik
Universität Augsburg
Memminger Straße 14
D-8900 Augsburg, FRG

Prof. Dr. Klaus Spremann
Abteilung Wirtschaftswissenschaften
Universität Ulm
Oberer Eselsberg N25
D-7900 Ulm, FRG

ISBN-13:978-3-540-51675-0 e-ISBN-13:978-3-642-75060-1
DOI: 10.1007/978-3-642-75060-1

2142/7130-543210

Preface

Agency Theory is a new branch of economics which focusses on the roles of information and of incentives when individuals cooperate with respect to the utilisation of resources. Basic approaches are coming from microeconomic theory as well as from risk analysis. Among the broad variety of applications are: the many designs of contractual arrangements, organizations, and institutions as well as the manifold aspects of the separation of ownership and control so fundamental for business finance.

After some twenty years of intensive research in the field of information economics it might be timely to present the most basic issues, questions, models, and applications. This volume *Agency Theory, Information, and Incentives* offers introductory surveys as well as results of individual research that seem to shape that field of information economics appropriately. Some 30 authors were invited to present their subjects in such a way that students could easily become acquainted with the main ideas of information economics. So the aim of *Agency Theory, Information, and Incentives* is to introduce students at an intermediate level and to accompany their work in classes on microeconomics, information economics, organization, management theory, and business finance.

The topics selected form the eight sections of the book:

1. Agency Theory and Risk Sharing
2. Information and Incentives
3. Capital Markets and Moral Hazard
4. Financial Contracting and Dividends
5. External Accounting and Auditing
6. Coordination in Groups
7. Property Rights and Fairness
8. Agency Costs.

More details are listed in the Table of Contents. The editors hope that the sequence of presentation permits an organic and sensible view of the whole topic.

Such a task could never be completed without the support and the advice given by other scholars and by anonymous referees. In addition, financial support was granted by Stiftung Volkswagenwerk and by the Landeszentralbanken in Bayern and in Baden-Württemberg. That permitted a scientific meeting of the contributors to be held at Schloß Reisensburg in the sum-

mer of 1986. Ideas, views and individual values could be exchanged among the scholars such that words, letters, and symbols communicated reflect, in some sense, ''aggregated perspectives'' of the subject under discussion.
Especially we want to extend our thanks to Birgit Emmrich for her patience during the different stages of manuscript preparation. Last but not least, we are indebted to Werner A. Müller from Springer Publishing Company for his readiness to present *Agency Theory, Information, and Incentives* with the same care he already published our volumes ''Risk and Capital'' (1984) and ''Capital Market Equilibria'' (1986).

Günter Bamberg
University of Augsburg

Klaus Spremann
Ulm University

Table of Contents

Section 3 Capital Markets and Moral Hazard

Section 4 Financial Contracting and Dividends

Section 5 External Accounting and Auditing

Section 6 Coordination in Groups

Section 8 Agency Costs

Prologue

Günter Bamberg and Klaus Spremann

If Economics can be correctly defined to be the science of cooperation with respect to the utilization of resources, economic analysis has to focus on arrangements, contracts, organizations, and institutions that set the rules according to which cooperation is taking place among individuals. Above all, such rules define ways of both co-ordination and participation, that is, the ways each individual is expected to contribute and how each individual participates and shares success or failure of the joint effort.

Although reality shows a broad variety if not a continuum of such arrangements, contracts, organizations, and institutions, it is useful to distinguish a few and characteristic types. Four such idealised types of organizations and institutional arrangements are represented through

- Competition (the decentralised co-ordination of markets)
- Regulation (enforcing contracts, centralised planning, control exercised by bureaucracies or government, penalties if rules are violated)
- Motivation (performance-oriented reward and incentives as they are common within private business enterprises)
- Socialization (the mutual adjustment, close observation, and help as provided in families and clans).

In all four types of organizations the scheme of cooperation has the same structure in so far as each individual is expected to give something and gets something in return.

To make sure, the market is the economists' favourite type of organizational design. Markets, however, will not work well in the presence of externalities. A second reason of market failure may be uncertainty about the qualities of the commodities, services, or rights exchanged. An important kind of uncertainty results from imperfect information, in particular, from asymmetric information.

Thus, whenever externalities and/or imperfect information prevail it might be necessary to replace pure competition by a mixtum compositum of competition and of the other three organizational designs: regulation, motivation, socialization. Since external effects and imperfect information are more the regular than the irregular case in real life, many economists extended the theory of pure competition into the directions indicated.

Agency Theory, in most general terms, can be viewed as the economic analysis of cooperation in situations where externalities, uncertainty, limited observability, or asymmetric information exclude the pure market organization. In fact, some scholars who focussed on incentive compatibility in the allocation of public goods (as an extreme case of external effects) meant to contribute to the theory of agency. The same is true for other researchers who analysed risk sharing in the presence of moral hazard. Likewise, those who studied the design of self selection schemes to induce individuals to reveal their utility function through choice, contributed to the theory of cooperation under asymmetric information.

Agency Theory, Information, and Incentives consequently covers a field of economic research much broader than the simple relation between two individuals, called principal and agent. Nevertheless, this principal-agent relation can serve as an elementary and basic cooperative unit. It is true that the economic theory of agency provides insights into the functioning of hierarchies, but is not restricted to these forms of cooperation. Though many important applications can be found in finance, Agency Theory deals with non-financial applications, too.

As always, many scholars contributed and have formed and constructed that field of economic knowledge. If it were to give reference only to a few selected scientists that expressed some of the major insights at an early stage of time, one could recall two papers that have nothing lost in their meaning and actuality:

ALCHIAN, ARMEN A. and HAROLD DEMSETZ: Production, Information Costs and Economic Organization. American Economic Review 62 (1972) 5, 777-795,

ARROW, KENNETH J.: The Limits of Organization. W.W. Norton, New York 1974.

Section 1 Agency Theory and Risk Sharing

Agent and Principal

Klaus Spremann

Summary:

In most general terms, agency theory focusses on cooperation in the presence of external effects as well as asymmetric information. To have a look on external effects first, consider two individuals. One of them, the agent, is decision making. He is thus affecting his own welfare and, in addition, that of the other individual called principal. These external effects of the agent's decisions or actions are negative: modifications of the agent's action which are preferred by the principal yield disutilities to the agent. A common example is a situation where the principal is assisted by the agent and the agent is deciding on level and kind of his effort. The principal is thus ready to pay some kind of reward to the agent in return for a certain decision/action/effort. Unfortunately, and this is the second characteristic of situations in agency theory, the principal cannot observe the agent's actions in full detail. The asymmetric information with respect to the agent's decision excludes simple agreements concerning pairs of action and payment.

External effects and asymmetric information prevail in very widespread situations of economic cooperation. The variety of examples include such important relations as those between employer and employee, stockholder and manager, or patient and physician.

From a methodological point of view, the principal-agent relation is closely related to risk sharing, hidden effort, monitoring, hidden characteristics, screening, and self selection. The purpose of this essay is to model and analyse these different features of agency theory in one unified approach. This formal approach is based on linear reward schemes, exponential utility functions, and normal distributions, and it will therefore be called LEN-Model. The LEN-Model allows for explicit presentation of endogenous parameters which determine the agent's decision on effort, the

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

chosen reward scheme, and the incorporation of monitoring signals. Hence several insights into how the pattern and design of cooperation depends on exogenous parameters such as the agent's risk aversion and the variance of environmental risk can be provided.

Table of Contents

1. A GENERAL VIEW

1.1 Cooperation

Economics may be viewed as the science of cooperation with regard to the utilization of resources. The basic pattern of cooperation is the exchange of goods, services, information, risk, or rights. If two or more individuals agree to cooperate, each of them will and has to contribute something and is going to receive something in return. Because of this pattern of exchange, the market is a very important organization or set of rules according to which cooperation takes place. Though the market mechanism is not the only design to organize cooperation, markets are efficient if the commodities exchanged have no external effects and if all relevant information is public.

More complex arrangements, however, are required in the presence of external effects or imperfect information. External effects prevail in such cases as that of non-separable labour inputs and that of public goods. Likewise imperfect information, in the sense of uncertainty about the quality of the commodities (skill and effort of labour input, reliability in financial contracting), require a more sophisticated design of the rules of cooperation.

Both external effects and imperfect information are predominating in many situations of economic cooperation. Usually these effects will be mutual. Each of the cooperating individuals affects by her/his decisions the welfare of the others directly, and each individual has some limits to observe the actions of others in full detail. Reciprocally given externalities and common limits to observe explain why cooperation is so complex in real life and why so many different types of arrangements, forms of contracts, institutions, and organizational designs have evolved.

Many approaches have been made to analyse the variety of arrangements. Among the first papers on agency theory are A.A. Alchian und H. Demsetz (1972), S.A. Ross (1973), J.E. Stiglitz (1974), M.C. Jensen and W.H. Meckling (1976). The economics of the principal-agent relationship were further developed, among others, by S. Shavell (1979), B. Holmström (1979, 1982), S.J. Grossman and O.D. Hart (1983). Recent surveys were presented by R. Rees (1985), by J.W. Pratt and R.J. Zeckhauser (1985), and by K.J. Arrow (1986). Many financial impacts of agency theory can be found in A. Barnea, R.A. Haugen and L.W. Senbet (1985).

1.2 External Effects

For analytical purposes one has to restrict the view on a simple, single-directed case of external effects and asymmetric information. So, instead of many, consider two individuals only. One of them, the agent, makes his decision $x \in X$. This decision, in some sense, is made on the quantity/quality of what the agent is going to contribute to what could be called the team. By this decision making the agent does not only influence his own welfare (more effort in team work is connected to individual disutility, for example) but also that of the other individual called principal. (The principal participates in the result of team work which is a consequence of the agent's effort). Agent and principal have different values associated with the agent's actions. In other words, the external effects of the agent's decision making are negative: those modifications of his action which are preferred by the principal yield disutilities to the agent.

Under such conditions, the principal is likely to start negotiations with the agent and offer some compensation, perhaps in form of a payment, if the agent refrains from choosing an action the principal dislikes. This way, both individuals could reach an agreement (x,p) that commits the agent to a certain decision $x \in X$ in exchange for a certain pay p to be made by the principal. It will be easy for them to arrive at an efficient agreement, which therefore could be termed first-best design of cooperation. Note that the agent's welfare or utility U(x,p) depends on pairs of action x and pay p (he prefers both lower levels of effort and higher payments). Likewise, the principal's welfare V(x,p) depends on pairs of action x and pay p (she prefers more effort of her partner as well as to give a lower pay). The situation of bargaining on pairs (x,p) can best be illustrated in an Edgeworth-Box.

1.3 Asymmetric Information

Externalities alone cause no deviation from first-best designs of cooperation. Simple bargaining on pairs of actions x and payments p are excluded, however, if external effects occur in combination with asymmetric information. Assume that, for some reason or the other (one reason is presented in Section 2.1) the principal is unable to observe and to verify exactly which action x the agent is or was realising. Information is asymmetric because the agent, of course, knows which decision he is going to make. But now, if there is no unlimited trust, it does not make sense for the principal to negotiate on pairs (x,p). The agent could make any promise with respect to his action and depart from it later on just because the principal is unable to control or to monitor the agent's decision making.

Although there is asymmetric information with respect to the agent's decision x by assumption, there might exist some variables which are correlated to x and the values of which can costlessly be observed by both agent and principal. Such variables provide some or partial information on the agent's decision x. Denote variables that partially inform on action x by y,z,... Depending on the particularities of the situation, examples for such variables are firstly the resulting output y of team work and secondly the monitoring signals z resulting from some control devices. Since the values of y and z can be observed by agent and principal without disagreement, reward schemes p(.,.) can be defined that make the amount of pay p(y,z) a function of these variables y,z. More details are presented in Section 2.5.

Now suppose the principal, unable to observe the agent's decision x in an exact and direct way, offers a certain reward scheme $p(.,.) \in P$, taken from a set P of feasible functions of variables y,z. The principal makes this offer without expecting any pretense or promise of the agent with respect to decision x. The principal just invites the agent to accept the scheme p(.,.) and to make, then, a decision x in his own interest. Consequently, there will be no shirking. The agent, realising that the actual pay p(y,z) depends on the values of the variables y,z which are related to his action x, will make his decision as a response to the scheme p. Formally, the agent is now choosing an action $x = \phi(p)$ that depends on the reward scheme p. The agent's response is described by the function $\phi: P \to X$. In other words, the reward scheme sets an incentive, or, the agent's decision x is induced by the reward scheme p.

1.4 Induced Decision Making

One consequence of information asymmetry is that only designs of cooperation are possible where the action $x = \phi(p)$ is induced by payment p. This is a fundamental difference between the first-best situation discussed in Section 1.2, where agent and principal could negotiate on pairs (x,p) of action and payment without further restriction. Under imperfect or asymmetric information, there is the additional constraint that the agent's action must be induced by payment.

Denote by E the set of pairs (x,p) that are efficient with respect to the welfare U of agent and the welfare V of principal. Thus E is the set of first-best designs of cooperation. Further, let I be the set of pairs (ϕ(p),p) of action and payment, where the action is induced by payment. The set I contains all designs that are feasible under information asymmetry. The information asymmetry would cause no problem at all if both sets E and I were identical. Any first-best design of cooperation could then be realized through induced decision making. One could already be satisfied in some weaker sense if the sets E and I had one or some elements in common. In such cases, at least one or some first-best designs of cooperation could be reached through induced decision making. Situations where E and I coincide or have some common elements are usually referred to as incentive compatibility.

In all other cases, the fact that some of the relevant information is not public causes a deviation from first-best and efficient designs (set E). Then all designs in I are dominated by designs in E and, for that reason, are second best only.

Few attempts have been made to measure the disadvantage between first-best and second-best designs in terms of a real number. Such measures are called agency costs in the tradition of M.C. Jensen and W.H. Meckling (1976). In figurative terms, agency costs measure the distance between the set E of first-best designs, which are an utopian fiction in the presence of asymmetric information, and the set I of designs where the agent's decision is induced by a payment scheme. The distance between two sets, however, can be measured in many different ways such that a particular definition of agency costs can easily be criticised with regard to appropriateness. In particular, one has to be very careful when using agency costs to compare and evaluate alternative second-best arrangements.

Another and presumable less ambiguous way is to define agency costs as the decision-theoretic value of perfect information: How much would the principal at most be willing to pay for becoming able to observe the agent's decision correctly? Agency costs as value of perfect information provide an upper bound for monitoring costs. If there were the possibility to introduce a perfectly working monitoring device it would be rejected if the costs of the device surmount the information value, see Section 2.4.

1.5 Hierarchy and Delegation

Note that no hierarchy was assumed so far. Neither was the principal assumed to be the boss nor the agent to be her subordinate as one might associate from the designations of the two cooperating partners. Consequently, the expression of a team seems to be much more appropriate. Agent is simply that member of the team who can vary his action/effort/behaviour/input. Principal is that member of the team who cannot costlessly observe the agent's action/effort/behaviour/input. Therefore, team members are bounded to schemes that set incentives. If person A buys insurance from company P, company P can hardly observe the care person A shows to avoid the accident, and nevertheless there is no hierarchical cooperation between A and P, see M. Spence and R. Zeckhauser (1971).

The relations between employer and employee as well as between stockholder and manager are very important examples for an agent-principal relation. Although most approaches are based on the identifications of principal and employer or principal and stockholder, resp., some aspects of these relations require to see the subordinate as principal and the superior as agent, see P. Swoboda (1987). In fact, the reward systems of hierarchical organizations sometimes provide more incentives for bosses than for subordinates.

Further, no formal contract was supposed to legalize the relation between agent and principal. Moreover, not necessarily it is the case that "the principal delegates some decision making to the agent", though the delegation of decision making provides a reasonable explanation of why the principal cannot observe the agent's doing in full detail. But there are many other situations different from the "delegation of decision making" where it is easy to see that the principal has some difficulties in controlling the agent's action/effort/behaviour/input. One example is the situation of insurance mentioned above.

1.6 Hidden Effort, Hidden Characteristics

The elaboration of Agency Theory requires a closer look to a number of different issues. One major task is to present a variety of different situations where a principal cannot completely observe an agent. In addition, reasonable argumentations have to be given for this information asymmetry. One should distinguish two situations which were termed by K.J. Arrow (1986): hidden efforts and hidden characteristics.

In many cases agent and principal cooperate within an organization and they know each other quite well. Each of them might provide some inputs to the team, but the principal's inputs are not under discussion here. The input provided by the agent are labour or management services and what can hardly be observed by others is the agent's effort. Effort is not only diligence and sweat but could also refer to the agent's renunciation of consumption on the job. Hidden effort and managerial discretion thus refer to the same situation.

The total team output and hence the principal's welfare depend on the agent's effort, but additionally also on some exogenous risk (state of nature). Although the principal knows the probability distribution of this risk, she might be unable to come to know which state nature was actually realizing. Consequently, she is unable to separate low effort from bad luck. If results turned out to be poor, the principal cannot conclude that the agent's effort must have been low. So it is the environmental uncertainty that explains why the principal is unable to deduce the agent's effort from the resulting team output.

As stated, the team members know each other. In particular, the principal knows the characteristics of her agent such as his skill and his attitude toward risk. Although the principal is unable to observe her agent's effort, she can predict the way in which the agent will behave under certain conditions. She can calculate the agent's response (function $\phi: P \to X$) to a certain reward scheme. The principal can thus study the impact of reward schemes on her own wealth, and, determine a reward scheme that is best with respect to her own interest and subject to the constraint that the agent's effort is induced by the reward scheme.

In the basic situation of hidden effort the reward will be a function of team output y. This can be generalized if there is a monitoring signal z, i.e., a statistic that is correlated to the agent's effort. The issue of <u>monitoring</u> is thus related to the situation of hidden effort.

A situation quite different from hidden effort is that of <u>hidden characteristics</u>. Here cooperation occurs across markets and the principal is unable to observe the agent's decision in time. A principal on the one side of the market gets into contact with many individuals, potential agents, on the other side. The principal has to make an offer in the moment of getting into contact with one of these agents. The agents, however, differ in their characteristics. Although the principal might know the distribution of characteristics, she usually will be uncertain about the particular type of agent. How to make an offer that is appropriate without knowing the individual characteristic?

In such cases of hidden characteristics the principal will look for <u>sorting devices</u> or install additional instruments that partially reveal hidden characteristics through <u>screening</u>. An important screening device consists of a set of payment schemes which allow for <u>self selection</u> through agents. Self selection schemes should be designed such that each agent has an incentive to reveal his type and his characteristics through choice. Such a scheme is presented in Section 2.6.

2. A CLOSER LOOK

2.1 Risk Sharing

A common situation of hidden effort is one in which the principal seeks help from the agent because her wealth depends on services the agent can provide. The agent can offer these services in various quantities and qualities upon which he alone decides. Formally, the agent chooses an element x from a set X of feasible actions. This decision, in its manifold aspects, is called effort. So far the external effects are outlined. On the other hand, the principal's wealth is not only affected by the agent's effort. Another factor is some kind of exogenous risk the probability distribution of which neither principal nor agent can control. Describe this state of nature by the random variable $\tilde{\theta}$. Thus the principal's gross wealth, denoted by $\tilde{y}$, can be viewed as a function of effort x and risk $\tilde{\theta}$,

$$\tilde{y} = f(x,\tilde{\theta}) . \qquad (1)$$

It might be indicated to visualize this situation as one of production although sometimes this notion must be interpreted in a broad sense. Anyway, the principal's gross wealth $\tilde{y}$ will be called <u>output</u> or <u>result</u>. The only <u>input</u> upon which a decision can be made is the agent's <u>effort</u> x. If there were any other inputs, their quantities and qualities will be supposed to be either fixed or settled beforehand.

Of course, the principal wants to buy some input from the agent but, unfortunately, she cannot observe how much the agent is providing and how good he is performing. In other words, the principal is assumed to be unable to observe the agent's effort decision $x \in X$. One implication of the exogenous risk $\tilde{\theta}$ is that it gives a reason for the assumed <u>information asymmetry</u>. If the principal is not completely ignorant, she will usually know the production function f (how her gross wealth is affected by her agent's effort and the exogenous risk), and she will know the probability distribution of $\tilde{\theta}$. Later she will also observe the realization y of her gross wealth $\tilde{y}$. But, to speak in figurative terms, she might be too distant from the location of production in order to see which state θ nature realized. Consequently, the principal cannot infer the agent's effort from the knowledge of both technology f and result y. The information asymmetry rules out negotiations with the aim to close with an agreement on effort.

Assume that the realization y of the output can be observed by both agent and principal correctly and without costs. Hence the principal can offer a <u>payment scheme</u> p(.) where the actual payment p(y) to be made to the agent depends on the realization y of output. Clearly, the principal will then keep the residuum y - p(y) as her net wealth. Denote by P the set of such schemes p(.) from which the principal is choosing one in order to offer it to her agent.

So far the agent need not make any committing declaration or contract in any legal sense. He will just realize the principal's offer, consider it in his decision-making calculations, and accept the money later when the realization y becomes known. Note, however, that for some reward scheme it could happen under a particular realization of output that the actual payment is negative. In such a case, the agent were to pay the corresponding amount to the principal. In order not to exclude such schemes from further consideration, the right will be assigned to the agent to decide whether or not to <u>accept</u> a payment scheme. If the agent accepts a payment scheme p(.) he declares himself willing to make an eventual transfer in the case p(y) is negative. But the agent is never supposed to make any promise with regard to his effort decision which could not be checked by the principal anyway.

Let c(x) be the agent's <u>disutility of effort</u> in terms of a money equivalent. So to speak, c(x) is the cost the agent has to pay by himself for the services he is going to provide as input. If the agent was offered and had accepted the payment scheme $p \in P$ and is now going to decide upon his effort $x \in X$, he is confronted with net wealth

$$(2) \qquad \tilde{w}(x,p) = p(f(x,\tilde{\theta})) - c(x) .$$

Since the result (1) is uncertain at that moment of decision making, the wealth w will be uncertain, too. In the particular case the scheme p(.) is constant in y such that the agent receives a fixed wage rather than sharing the result, his wealth is free of risk. The welfare derived from wealth w can be formalized by the expected utility $E(u(\tilde{w}))$, or, what is done here, the agent's welfare U is expressed in terms of the certainty equivalent

$$(3) \qquad U(x,p) := u^{-1}(E[u(\tilde{w})]) .$$

Thereby, u denotes the Neumann-Morgenstern utility function of the agent. He is supposed to be risk averse (u is concave), and hence the certainty equivalent U of wealth is below the expected value $E[\tilde{w}]$. The difference between the two entities was called <u>risk premium</u> by J.W. Pratt (1964).

A second implication of the exogenous uncertainty $\tilde{\theta}$ introduced in (1) is that it raises the issue of risk sharing. The more a payment scheme lets the agent share the uncertain result $\tilde{y}$, the more risky becomes his wealth (2). Suppose the principal wants to set an incentive to her agent by offering a considerable result sharing. The agent is not only requiring a compensation for his disutility of effort c(x). Because of his risk aversion, the agent needs also a higher risk premium in order to maintain a certain level of welfare.

That risk premium may turn out to be inefficient from a risk-sharing point of view. Suppose the principal is risk neutral so she could bear all the risk without requiring a premium. The principal keeps all the risk with her residuum $\tilde{y} - p(\tilde{y})$ if the scheme p(.) is constant such that the agent receives a fixed wage independent of the uncertain result. Such a fixed-wage payment, however, will set no incentives.

2.2 Induced Effort

How will the agent respond to a payment scheme p(.)? He will choose his effort such that his welfare (3) is maximized. Let $x^* \in X$ denote an optimal decision,

$$(4) \qquad U(x^*,p) \;=\; \max \{U(x,p) \mid x \in X\} \;.$$

The effort choosen depends, among other things, on the payment scheme and hence we write $x^* = \phi(p)$. Omit questions of existence and uniqueness (for some of the problems involved see S.J. Grossman and O.D. Hart (1983)), and solve (4) for each $p \in P$. This yields the <u>response function</u> $\phi: P \rightarrow X$ that describes the way in which the agent responds to reward schemes. In other words, ϕ describes how effort is induced. Note that under scheme p the agent can and will attain the welfare $U(\phi(p),p)$.

The decision on effort is not the only choice to be made. Distinguish four consequential choices. The <u>first</u> choice is made by the principal who selects a payment scheme $p \in P$ and suggests it to the agent. The second decision is made by the agent when he either accepts or refuses the scheme suggested. The agent makes his decision on acceptance in view of some other opportunities he might have and the best of which guarantees a certain reservation welfare m. Evidently, the agent is accepting a payment scheme p only if the welfare attained is not below the reservation level.

$$(5) \qquad U(\phi(p),p) \geqq m \ .$$

For that reason, the inequality (5) is called reservation constraint. If the agent refuses, the principal will presumably suggest another payment scheme. So there might be some bargaining and the first two decisions turn out to be interrelated. To make here a clear statement, we proceed on the assumption that the agent accepts a scheme p if and only if the reservation constraint (5) is satisfied. The reservation level m is thereby either belonging to the data or is resulting from negotiations. In short, m is considered as an exogenous parameter.

The third decision: If the agent accepted a reward scheme p he is going to choose his effort $x^* = \phi(p)$. The fourth and final step of that sequence is the realization of the state of nature, more precisely, the realization y of $\tilde{y}$ becomes known to both principal and agent. Only now the actual payment p(y) can be made. This ends the cooperation.

Nothing was said hitherto about the first decision in that chain of four choices. How will the principal choose a scheme p from set P? The principal's wealth is the residuum $\tilde{y} - p(\tilde{y})$, and her welfare (again expressed in terms of a certainty equivalent) is

$$(6) \qquad V(x,p) \ = \ v^{-1}(E[v(\tilde{y} - p(\tilde{y}))]) \ ,$$

where v denotes the Neumann-Morgenstern utility function of the principal. The welfare (6) depends on the agent's effort x since the result $\tilde{y}$ depends on x.

One of the stronger assumptions in the hidden-effort situation is that the principal knows all relevant characteristics of the cooperating agent. The relevant characteristics of the agent are: utility function u, disutility c(.), set of feasible effort decision X, and the reservation level m. With that knowledge the principal can calculate the way ϕ in which the agent will respond

$x^* = \phi(p)$ to reward schemes $p \in P$. This assumption simplifies the principal's decision to

(7) maximize $V(\phi(p),p)$ with respect to $p \in P$
subject to the reservation constraint (5).

A solution of (7) will be denoted by p_m^*. As was indicated by the subcript m, the reservation level usually has a major impact on the scheme selected. Of course, the optimal scheme also depends on data such as the technology f, the agent's risk aversion $-u''/u'$, and the variance $Var(\tilde{\theta})$ of the exogenous risk.

A final remark is made on the assumption according which the principal knows the agent's characteristics and is thus in the position to predict her agent's decision making although she is, due to the information asymmetry, unable to verify her calculations by observation. What makes then the difference between the ability to predict and the ability to observe? Suppose the principal selects the scheme p and predicts, by herself, that the agent will respond with effort $x^* = \phi(p)$. What the agent will do in fact is to choose exactly that effort x^*. The problem is not that there could be any difference between what the principal predicts and what the agent really does. The principal's prediction is always correct.

Rather than that the true problem is: both individuals cannot freely negotiate in order to agree upon any pair (x,p) of effort and payment. Suppose, for a moment, both individuals would agree to realize a particular pair $(\bar{x},\bar{p})$ where $\bar{x} \neq \phi(\bar{p})$. Then the principal, unable to observe the agent, can predict that the agent will realize the effort $x^* = \phi(\bar{p})$ in disaccord with the agreement. And the selfish agent will, in fact, make his decision x^* as predicted. Consequently, both individuals are restricted in their cooperation to those specific pairs (x^*,p), where effort is induced by the payment $x^* = \phi(p)$. For that reason, there is no need and no sense to discuss on effort at all. Agent and principal just speak on payment schemes p and none of them has doubts about the corresponding effort induced. Since they do not settle effort, there is no shirking.

The discussion between agent and principal on the payment scheme was modelled here in that way: The principal selects, from all payment schemes which guarantee the agent a certain welfare $U \geqq m$, that scheme p_m^* which maximizes her own welfare V. The resulting design of cooperation is characterised by the pair (x_m^*, p_m^*) of induced effort $x_m^* = \phi(p_m^*)$ and payment p_m^*. By variation of the parameter m one gets the elements of the set I of second-best designs defined in Section 1.4.

2.3 The LEN-Model

The hidden-effort situation as outlined in the last section cannot be solved in its general form. In order to study how the induced effort and the selected payment scheme depend on the data and parameters of the model, we further specify functions and variables. The set of specifying assumptions suggested here is called Linear-Exponential-Normal-Model, since

(L) output $\tilde{y}$ is a linear function of risk $\tilde{\theta}$, and feasible payment schemes $p(\cdot) \in P$ are linear functions of output,

(E) the utility function u of the agent is exponential; likewise the principal has constant absolute risk aversion,

(N) the risk $\tilde{\theta}$ is normally distributed.

Specifications (N), (L) imply that both the agent's wealth and the principal's residuum are normally distributed. That, in conjunction with (E), implies that the certainty equivalents (3), (6) can be expressed as expected value minus half the variance times risk aversion (G. Bamberg and K. Spremann (1981)). A simple version of the Len-Modell is:

$$\tilde{y} = f(x,\tilde{\theta}) := x + \tilde{\theta} \qquad x \in X := [0,1/2]$$

$$\tilde{\theta} \text{ normal}, \quad E[\tilde{\theta}] = 0, \quad Var[\tilde{\theta}] = \sigma^2$$

$$p \in P \text{ if and only if } p(y) = r + sy$$

$$u(w) = -\exp(-\alpha w), \ \alpha > 0$$

v linear (principal is risk neutral)

$$c(x) = x^2 .$$

The agent's effort has one dimension only and the result $\tilde{y}$ is the sum of effort x and the one-dimensional random variable $\tilde{\theta}$. The agent has constant risk aversion denoted by $\alpha = -u''/u' > 0$ and the principal is risk neutral $-v''/v' = 0$. In order to describe increasing marginal disutility of effort, the function c(.) is supposed to be quadratic.

Two parameters r,s determine feasible payment schemes: r will be called <u>fee</u> and s will be called <u>share</u>. So far there are no restrictions on $r,s \in \mathbb{R}$ although the share may be viewed as constrained to $0 \leq s \leq 1$. In the case s = 0 the principal pays a fixed fee r for the services provided by the agent, independent of team profit. In the case s = 1 it is the agent who bears all the risk, while the principal's wealth will be risk free under such an agreement. The fee r can be negative, too, which could be indicated in particular if the agent receives a positive share s > 0 of the result. One could then refer to r as a <u>rent</u> paid to the principal, and we will use the term rent independent of whether r is positive, negative, or equal to zero. From now on, we write the scheme as pair (r,s) of rent and share.

Analysis and results presented in the sequel depend, as always, on the specific assumptions made. In particular, the class of linear payment schemes has a major impact. We just mention that non-linear schemes have been suggested. Quite often arrangements can be found where the agent's reward is not a linear function of team output. Sometimes, the agent participates in gains but not in losses.

(8) $\quad p(y) = r + s \cdot \max\{0,y\}$

such that the risk premium demanded will be reduced. Denote the set of payment schemes (8) by P_+. Not only is the question which are the parameters r,s chosen in the situation where P_+ is the set of feasible arrangements. Another issue is whether or not the best schemes in P_+ are superior to linear profit-sharing arrangements in P. In other words: which arrangements would be chosen in the set $P \cup P_+$? Another common arrangement is a bonus-penalty scheme: a fixed fee r is applied as long as the profit is not below a certain critical level y_L, combined with a fine $t \geq 0$ for too poor results:

(9) $$p(y) = \begin{cases} r & \text{if } y \geqq y_L , \\ r-t & \text{if } y < y_L . \end{cases}$$

For an analysis of penalty schemes see also J. Mirrlees (1975).

Our version of the LEN-Model has three <u>exogenous</u> parameters: the agent's risk aversion $\alpha > 0$, the reservation level m, and the variance $\sigma^2 > 0$ of the environmental risk. There are three <u>endogenous</u> variables: the agent's effort x, and rent r and share s which determine the payment scheme. The purpose is to study how the induced effort and how the payment scheme (r,s) depend on the exogenous parameters.

The principal will find an optimal payment scheme (r_m^*, s_m^*) in three steps which answer three questions. The <u>first question</u> is: in which way will the agent respond x* to a payment scheme (r,s)?

In the LEN-Model the agent's wealth (2) is equal to

(10) $$\tilde{w}(x;r,s) = r + (x + \tilde{\theta})s - x^2$$

and the derived welfare (certainty equivalent (3)) is

(11) $$\begin{aligned} U(x;r,s) &= E[\tilde{w}] - \frac{\alpha}{2} \operatorname{Var}[\tilde{w}] = \\ &= r + sx - x^2 - \frac{\alpha}{2} s^2 \sigma^2 . \end{aligned}$$

Maximization of U with respect to effort x yields the agent's response

(12) $$x^* = \phi(r,s) = \frac{s}{2} .$$

Note that $0 \leqq s \leqq 1$ implies $x^* \in X$. This proves:

THEOREM 1: Neither the rent r nor (the result of negotiations on) the reservation level m have an impact on the agent's effort. In particular, a fixed-fee arrangement, s = 0, induces the agent to the lowest feasible effort, $x^* = \phi(r,0) = 0$, however large the rent r may be.

The <u>second question</u> is: which payment schemes (r,s) will be accepted by the agent in view of the reservation constraint (5)?

Equations (11), (12) imply that the attained welfare is

(13) $$U(x^*;r,s) = r + \frac{s^2}{4}(1 - 2\alpha\sigma^2)$$

such that the reservation constraint is satisfied if the rent r has the size

(14) $$r = m - \frac{s^2}{4}(1 - 2\alpha\sigma^2) ,$$

at least.

A common hypothesis is that the fee or rent r can be reduced in an arrangement if the share s is increased. As (14) indicates, however, that is correct only if both the agent's risk aversion α and the variance σ^2 are small enough, i.e., if $2\alpha\sigma^2 < 1$. To see the reason, recognize the difference between expected value and certainty equivalent of the agent's wealth (10) as a risk premium. The risk premium is equal to $(\alpha/2)\ s^2\sigma^2$. Thus a rising share s has three effects. (i) A higher bonus simply increases the expected income. (ii) A higher share induces the agent to more effort and he is participating in a better result. (iii) The agent is demanding a higher risk premium because he is going to bear more of the risk as the share is increased. If the third effect outweighs the first and the second effect, the overall result is that the fee r has to be increased instead of decreased as a higher share is envisaged.

THEOREM 2: If the agent's risk aversion α and/or the variance $\sigma^2$of the environmental risk are lage (in the sense of $1 < 2\alpha\sigma^2$), an increase of the share s requires an increase of the rent r.

The rationale of this result is that the agent will not only share in "profits" $\tilde{y} > 0$ but in "losses" $\tilde{y} < 0$, too.

The third question is: which payment scheme (r,s) maximizes the principal's welfare given the agent's response (12) and subject to the reservation constraint(in the form (14))?

The principal's wealth is the residuum

(15) $$\tilde{y} - (r + s\tilde{y}) = (1 - s)(x + \tilde{\theta}) - r$$

and her welfare, because of her risk neutrality, is the expected wealth

(16) $$V(x;r,s) = (1 - s)\ x - r.$$

Considering induced effort (12), the principal wants to maximize

$$(17) \qquad V(x^*;r,s) = (1 - s)\frac{s}{2} - r$$

with respect to rent r and share s such that the reservation constraint (14) is satisfied. Insert (14) into (17) and see that it means to maximize

$$(18) \qquad V = (1 - s)\frac{s}{2} - m + \frac{s^2}{4}(1 - 2\alpha\sigma^2)$$

which turns out to be a function of s alone. The share that maximizes (18) is easily determined,

$$(19) \qquad s_m^* = \frac{1}{1 + 2\alpha\sigma^2} .$$

From (14), (19) follows the rent selected,

$$(20) \qquad r_m^* = m - \frac{1 - 2\alpha\sigma^2}{4(1 + 2\alpha\sigma^2)^2}$$

whereas (12) gives the induced effort

$$(21) \qquad x_m^* = \phi(r_m^*, s_m^*) = \frac{1}{2(1 + 2\alpha\sigma^2)} .$$

Remember that the agent's welfare is $U^* = m$ whereas the principal attains the welfare

$$(22) \qquad V^* = \frac{1}{4(1 + 2\alpha\sigma^2)} - m .$$

The agent's share s_m^* effort x_m^*, expected output $E[f(x_m^*,\tilde{\theta})]$, and the principal's welfare V^* are inversely related to $\alpha\sigma^2$.

One implication of (20) is that the principal's welfare is inversely related to her agent's risk aversion α.

THEOREM 3: If the principal could choose between two agents who differ only with respect to risk aversion, she prefers the agent with the lower risk aversion.

To comment on (19), the principal finds it best to reduce the share s^* as the agent's risk aversion α and/or the variance σ^2 of exogenous risk increase. This is because the agent will then ask a higher risk premium. However, the agent will never get a fixed-fee salary.

THEOREM 4: No fixed-fee agreement (r,0) will be made, however large the agent's risk aversion is.

On the other hand, the principal prefers to keep a residuum almost free of risk, $s^* \approx 1$, if the agent's risk aversion α or the variance σ^2 are small. In such situations it is cheap to motivate through profit sharing since the risk premium required by the agent is small. An extreme situation is that of a risk neutral agent $\alpha = 0$. A risk neutral agent bears all the risk, $s_m^* = 1$. The effort $x_m^* = 1/2$ induced by that share $s_m^* = 1$ can be seen as first best as will be shown in the next section. The welfare attained by the principal assumes the largest value ever possible $V^* = 1/4 - m$. One can therefore conclude:

THEOREM 5: It is the connection of unobservability (of the agent's effort) and of risk aversion (of the agent) that excludes first-best arrangements.

2.4 Agency Costs

In the seminal paper by M.C. Jensen and W.H. Meckling (1976) agency costs were proposed to be a key tool in evaluating alternative designs of a principal-agent relation. The authors defined agency costs as the sum of (i) the monitoring expenditure by the principal (no such expenditures are modelled here), (ii) the bonding expenditures by the agent, and (iii) the residual loss, i.e. the monetary equivalent of the reduction in welfare experienced by the pricipal due to the divergence between the agent's decisions and "those decisions which would maximize the welfare of the principal" (1976, p. 308). The latter formulation, however, is not clear and ambiguous if taken literally, see D. Schneider (1987), R.H. Schmidt (1987).

In Section 1.4 agency costs were defined as an index that measures the distance between the set E of first-best designs and the set I of second-best designs. In order to determine agency costs along this line, we have to specify what measure of distance between E and I should be used. In addition, one has to explore for what purposes that index termed agency costs can serve. Agency costs as measure of distance can be presumed to give an estimation of how much the given second-best design could be improved if there were a monitoring device informing on the agent's effort. In fact, the nature of agency costs will be seen as a decision-theoretic value of perfect information. But one should be very careful when agency costs are suggested as a tool to evaluate alternative second-best designs.

Formally, we consider two particular designs, one belonging to the set E, the other to I. Both designs assign the same level m of welfare to the agent. Agency costs, in the sense of a distance measure, are the difference of the principal's welfare in these two designs. In the LEN-Model it turns out that this difference is independent of the parameter m. This rather abstract definition will now be made more concrete in terms of the information value.

The rationale of the principal-agent relationship is that the agent's effort cannot be observed by the principal. A rigorous approach has thus to define agency costs as a value of information: how much will the principal offer, at most, if he could observe the agent's effort?

If the principal has perfect information on the true effort of the agent, both team members can bargain and agree upon any effort in exchange for any payment. No longer has effort to be induced by a payment scheme. Under perfect information the principal would thus address to

$$\begin{aligned} &\text{Maximize} \quad V(x,p) \\ (23) \qquad &\text{subject to} \quad U(x,p) \geqq m \\ &\text{with respect to} \quad p \in P \text{ and } x \in X \end{aligned}$$

if the agent is willing to enter into cooperation as long as his welfare reaches the level m. Denote a solution of problem (23) by (x_m^o, p_m^o). This design (x_m^o, p_m^o) is the first-best design chosen to represent E.

From the set I we choose the design (x_m^*, p_m^*) that solves the problem

$$(24)\quad \begin{array}{ll} \text{Maximize} & V(x,p) \\ \text{subject to} & U(x,p) \geqq m \\ \text{with respect to} & p \in P \\ \text{where } x = \phi(p) \text{ is induced.} & \end{array}$$

Agency costs are now defined as difference

$$(25)\qquad AC_m := V(x_m^o, p_m^o) - V(x_m^*, p_m^*) .$$

For the LEN-Model it is easy to see that

$$(26)\qquad x_m^o = 1/2 .$$

This is the efficient effort upon agent and principal would agree if effort could be observed. The payment they will agree upon is

$$(27)\qquad \begin{aligned} r_m^o &= m , \\ s_m^o &= 0 . \end{aligned}$$

The results (26), (27) and (22) yield agency costs in the LEN-Model as

$$(28)\qquad AC = \frac{\alpha\sigma^2}{4\alpha\sigma^2 + 2} ,$$

independent of the agent's welfare m.

Since AC is increasing with $\alpha\sigma^2$ we get

THEOREM 6: The unobservability of the agent's effort becomes as more a drawback the larger the agent's risk aversion and the larger the variance of the environmental risk are.

Another insight provided by (28) concerns the output variance σ^2. The theory of finance tells that diversification is not an issue for the single firm because all unsystematic risks can be eliminated in well diversified portfolios. This result, however, remains no longer true in the context of agency theory. Diversification within a single firm implies a lower variance σ^2 and thus reduced agency costs. Consequently, the dependency of agency cost on the variance σ^2 may suggest to form teams, where the team output has, because of diversification, lower variance compared with the output variances of separated units. The reward of team members is then made as depending on the output of the whole team, rather than making reward a function of individual output.

The literature on teams, see A.A. Alchian and H. Demsetz (1972), often presents this rationale for the existence of teams: the team output can be observed but not be separated and presented as sum of what each of the team members contributed. The analysis presented here suggests another rationale for the existence of teams: the team output is diversified (lower σ^2) and hence, taken as a basis to reward team members, reduces the required risk premiums.

A final implication of (28) concerns the question of what happens in cases where the variance σ^2 is quite large and cannot be reduced through diversification. It may thus happen that agency costs are so high or, equivalently, that the principal's welfare (22) is such low that she prefers no cooperation with an agent at all. The principal, perhaps, has other opportunities which determine a certain reservation level also for herself. Three ways to overcome such a situation of too high agency costs can be outlined.

Firstly, one could enlarge the set P of feasible payment schemes. Consider nonlinear schemes of the form (8) or bonus-penalty schemes of the form (9). If such schemes were feasible it could be the case that second-best designs come closer to first-best results. In a particular setup J. Mirrlees (1975) demonstrated the superiority of payment schemes that impose heavy penalties on a suitable range of outputs.

Secondly, one could consider monitoring devices that give additional information, though not perfect in every case, on the agent's effort. Costless monitoring signals were introduced by M. Harris and A. Raviv (1979), B. Holmström (1979), S. Shavell (1979), F. Gjesdal (1982), N. Singh (1985). An analysis of monitoring signals within the LEN-Model follows in the next section, and extension of these results with respect to costly monitoring was done by M. Blickle (1987).

Thirdly, society could encourage trust. If nowhere cooperation is starting, society can be supposed to develop and to reward behavior such as honesty, reliability, and altruism. In the literature on organization, such forms of behavior are induced through the process of indoctrination: the member of the organization internalizes cooperative criteria, which replace selfishness even within the reign of managerial discretion.

2.5 Monitoring Signals

Both the general model (Sections 2.1, 2.2) and the special LEN-Model (Section 2.3) on the hidden-effort situation can be extended in order to incorporate organizational instruments which monitor the agent and measure his effort. Generally speaking, there might be a (multidimensional) signal $\tilde{z}$ which, more or less exactly, reveals the agent's (multidimensional) effort x. Such a monitoring signal $\tilde{z}$ must thus be a function of x, though not a function of x alone. More or less exactly means that some additional uncertainty influences the value z of the signal $\tilde{z}$,

$$(29) \qquad \tilde{z} = h(x) + \tilde{\varepsilon} .$$

Such a monitoring signal may be seen as a sufficient statistic.

Both principal and agent are supposed to know the <u>observation function</u> h as well as the probability distribution of the <u>observation error</u> $\tilde{\varepsilon}$. Nature will realize the random variable $\tilde{\varepsilon}$ at the same time when $\tilde{\theta}$ is realized. Like $\tilde{\theta}$, the principal will not learn the realization ε of $\tilde{\varepsilon}$. Nevertheless, the principal can find it better to make the reward not only depending on output y but also on the monitoring signal z. When cooperation is started, the principal is thus suggesting a payment scheme p(y,z) as a function of output y and the monitoring signal z.

In its simplest form, the observation function h and the parameters of the probability distribution of the observation error $\tilde{\varepsilon}$ are given beforehand. The principal has thus to decide whether to utilize the signal z in the reward scheme or not, and if yes, in which way the payment should depend on z. Such an extension will now be studied within the framework of the LEN-Model.

In more complex cases, the form of the observation function h or distributional parameters of the observation error $\tilde{\varepsilon}$ might belong to the principal's decisions. Even more, alternative monitoring devices can imply different monitoring cost. To give an example, let effort x and signal z be one-dimensional variables, $z = x + \tilde{\varepsilon}$, $E[\tilde{\varepsilon}] = 0$, and the variance $Var[\tilde{\varepsilon}] = \sigma_\varepsilon^2$ being a decision variable. Thereby, monitoring costs increase in some way as a smaller error variance is chosen. The question is then not only how to make the reward depending on z but also: how much wealth should be devoted to make monitoring more precise, i.e., to reduce σ_ε^2.

Now, a costless monitoring signal is considered and introduced into the LEN framework. A straightforward extension of the simple version of the LEN-Model presented in Section 2.3 is

output $\tilde{y} = f(x,\tilde{\theta}) = x + \tilde{\theta}$,

signal $\tilde{z} = h(x) + \tilde{\varepsilon} = x + \tilde{\varepsilon}$,

$\tilde{\theta} \approx N(0,\sigma_\theta^2)$, $\tilde{\varepsilon} \approx N(0,\sigma_\varepsilon^2)$, $Cov(\tilde{\theta},\tilde{\varepsilon}) = 0$,

effort $x \in X = [0,1/2]$

payment scheme $p \in P$ iff $p(y,z) = r + sy + tz$,

agent's risk aversion $-u''/u' = \alpha$,

principal risk neutral ,

disutility of effort $c(x) = x^2$.

The analytical solution follows the steps presented in Section 2.3. The principal wants to choose a reward scheme given by the triple (r,s,t) such that her welfare, the expected resiudal wealth $\tilde{y} - (r + s\tilde{y} + t\tilde{z})$, is maximized. As before, the principal knows the agent's characteristics α,X,c,m and can thus predict the agent's response $x^* = \phi(r,s,t)$ to a reward scheme (r,s,t).

The agent's wealth, similar to (10), is normally distributed,

$$(30) \qquad \tilde{w}(x;r,s,t) = r + s(x + \tilde{\theta}) + t(x + \tilde{\varepsilon}) - x^2$$

and the welfare (certainty equivalent) is equal to

$$(31) \qquad U(x;r,s,t) = r + (s + t)x - x^2 - \frac{\alpha}{2}(s^2\sigma_\theta^2 + t^2\sigma_\varepsilon^2) .$$

Maximization of (31) with respect to x yields

$$x^* = \phi(r,s,t) = \frac{s+t}{2}, \tag{32}$$

which is the induced effort.

Further, the agent is assumed to accept a reward scheme (r,s,t) if and only if the reservation constraint $U(x^*,r,s,t) \geqq m$ is satisfied. A remark on the reservation level will be made below. The constraint requires a rent r which has the level

$$r = m + \frac{s^2}{4}(2\alpha\sigma_\theta^2 - 1) + \frac{t^2}{4}(2\alpha\sigma_\varepsilon^2 - 1) - \frac{st}{2} \tag{33}$$

at least. Given the response (32) and the reservation constraint (33), the principal's welfare is

$$\begin{aligned} V &= E[\tilde{y} - (r + s\tilde{y} + t\tilde{z})] = \\ &= \frac{s+t}{2} - m - \frac{s^2}{4}(2\alpha\sigma_\theta^2 - 1) - \frac{t^2}{4}(2\alpha\sigma_\varepsilon^2 - 1) + \\ &\quad + \frac{st}{2} - \frac{(s+t)^2}{2} . \end{aligned} \tag{34}$$

This welfare, taken as a function of s and of t is, for α sufficient small, concave and will be maximized for

$$s^* = \frac{1 - t^*}{1 + 2\alpha\sigma_\theta^2}, \qquad t^* = \frac{1 - s^*}{1 + 2\alpha\sigma_\varepsilon^2} . \tag{35}$$

The linear system (35) has the explicit solution

$$s^* = \frac{1}{1 + 2\alpha\sigma_\theta^2 + \sigma_\theta^2/\sigma_\varepsilon^2}, \qquad t^* = \frac{1}{1 + 2\alpha\sigma_\varepsilon^2 + \sigma_\varepsilon^2/\sigma_\theta^2} . \tag{36}$$

Equations (36), together with (33), provide the reward scheme (r*,s*,t*) selected. Some of the properties are noteworthy.

At first $t^* > 0$, in particular, $t^* \neq 0$. This means that the principal prefers to make the reward depending on the monitoring signal $\tilde{z}$ however inaccurate it is, i.e., however large the variance σ_ε^2 of the observation error may be. If that variance becomes larger and larger, ceteris paribus, t^* is chosen smaller, and the share s^* selected increases and tends to the value (19).

Another result of (36) concerns the question whether a wage should be paid for labor input (time of presence) or for labor output (result of work). Consider again a varying exactness of the monitoring signal as measured by the variance σ_ε^2. As σ_ε^2 becomes smaller, the share s^* decreases. If σ_ε^2 tends to zero, which means that the effort (labour input) can almost accurately be observed, the optimal share vanishes, $s^* = 0$. At the same time, t^* tends to 1 where 1 is the marginal value of effort to the principal. If effort were observable, the optimal reward scheme is just a price for units of effort, and each effort unit is rewarded according to its marginal value. The rent r has a distributional effect only. So understand the share s as a <u>wage</u> paid <u>for labour output</u> and t as a <u>wage for labour input</u>. The formula (36) indicates how to mix a payment for output with a payment for input in cases of observation errors.

A further comment is made on the bias to signal effort instead of really working. Extend the LEN-Model such that the agent's effort $x = (x_1, x_2)$ is now a two-dimensional decision variable. Symmetry is achieved through assumptions

$$\text{(37)} \qquad \begin{aligned} &\text{output} \quad \tilde{y} = f(x,\tilde{\theta}) = x_1 + x_2 + \tilde{\theta}, \\ &\text{disutility} \quad c(x) = x_1^2 + x_2^2 \,. \end{aligned}$$

This makes all efforts (x_1, x_2) with $x_1 \neq x_2$ inefficient, since the effort $\left(\frac{x_1 + x_2}{2}, \frac{x_1 + x_2}{2}\right)$ yields the same result at reduced disutility: $x_1^2 + x_2^2 > 2((x_1 + x_2)/2)^2$.

Now assume there is a monitoring signal which informs on x_1 but not on x_2,

$$\text{(38)} \qquad \text{signal } \tilde{z} = x_1 + \tilde{\varepsilon} \,.$$

The principal will then choose a reward scheme were the agent's salary depends on the monitoring signal, $t^* \neq 0$, although the agent will respond with an inefficient action $x_1 > x_2$. So to speak, the principal is aware and predicts that the agent utilizes working time to signal effort rather than to work. Nevertheless, the principal prefers to have the signal be part of the reward.

A final remark concerns the question whether or not the agent refuses cooperation when the principal is going to introduce an additional monitoring device. As an homo economicus, the agent would be indifferent if his welfare was unchanged. Might be the principal is willing to increase that reservation level m and is, nevertheless, better off. A principal prepared to modify m can expect that the agent looks by himself for signals that inform on his effort. Another point is that some kinds of monitoring devices cause additional disutilities to the agent which need compensation. Consequently, there are three reasons why the introduction of monitoring signals can be costly. One of course is that the technology of monitoring requires resources. The second reason is that the agent asks for new negotiations on the level m of his welfare. The third reason is that a disutility caused by monitoring must be compensated.

2.6 Screening

So far, the principal has been assumed to know all the agent's characteristics such as risk aversion and so forth. The analysis presented in Sections 2.1 through 2.5 focussed on hidden effort and monitoring. As outlined in 1.6, another issue of agency theory is screening. Sorting and screening devices become necessary in situations where some of the agent's characteristics are hidden. Perhaps the most important class of screening devices is that of self selection schemes. This is because everybody prefers free choice to inquisition even if the final outcomes are the same.

This section presents basic ideas in the design of self-selection schemes, see K.J. Arrow (1986). To be designed is a set of contracts such that each individual chooses the contract which is designed to fit his or her type. Thus, the individual's characteristics are revealed through choice.

Here, the hidden characteristic is supposed to be the agent's risk aversion. All results are derived within the framework of the LEN-Model. But no monitoring signals are considered in order not to burden the notation.

Consider a labour market with job searchers (possible agents) on the one side and the principal in search of an agent on the other side. This time, the issue is cooperation happening across the market. The principal is ready to offer reward schemes (r,s) with

(39) share $s(\alpha) = \dfrac{1}{1 + 2\alpha\sigma^2}$

(40) rent $r(\alpha,m,s) = m - \dfrac{s^2}{4}(1 - 2\alpha\sigma^2)$,

see (19), (14). However, the principal does not know a job searcher's risk aversion α this time.

Note that (39) is the optimal share as a function of risk aversion α, whereas (40) denotes the smallest rent an agent with risk aversion α and reservation level m would accept, if the share were equal to s, independent of whether or not this s is optimal in the sense of (39). Another result that should be recalled is the certainty equivalent of an agent's wealth under contract (r,s), now denoted by $U_\alpha(r,s)$:

(41) $U_\alpha(r,s) = r + \dfrac{s^2}{4}(1 - 2\alpha\sigma^2)$,

see (13). Suppose the principal knows there are two types k = 0,1 of job searching agents who differ only with respect to their risk aversion α_0,α_1. The principal knows further that low-risk-averse job searchers k = 0 are, to simplify notation, risk neutral and that type-1 job searchers have risk aversion α,

(42) $\alpha_0 = 0$ and $\alpha_1 = \alpha$.

Both types of agents may have, so far, the same exogenous reservation level m and identical disutilities of effort. As has already been demonstrated, the principal prefers type 0 job searchers to type 1 agents. According to (22), the principal's difference in welfare of getting a type-1 or type-0 agent is

$$(43) \qquad V_0 - V_\alpha = \frac{1}{4} - \frac{1}{4(1 + 2\alpha\sigma^2} = \frac{\alpha\sigma^2}{2(1 + 2\alpha\sigma^2)} .$$

But, in order not to conclude with a search model, assume the principal's aim is not to refuse type 1 agents. She just wants to offer to each job searcher a contract which she, the principal, finds best.

Everything were easy if the principal could costlessly find out a job searcher's type. He would then offer

contract 0 :

share $s_0 := s(0) = 1$

rent $r_0 := r(0,m,s_0) = m - \frac{1}{4}$

to type 0 agents, and

contract 1 :

share $s_1 := s(\alpha) = \frac{1}{1 + 2\alpha\sigma^2}$

rent $r_1 := r(\alpha,m,s_1)$ according to (40)

to individuals type 1.

What will happen if the principal cannot identify an agent's type and is going to allow all job searchers to choose among contracts 0,1?

The answer is that <u>the set of contracts</u> 0,1 <u>breaks down as a self-selection device</u>: Agents of any type decide for contract 1. The proof is twofold. Firstly agents type 1 understand that

$$(44) \qquad U_\alpha(r_0,s_0) = m - \frac{\alpha\sigma^2}{2} < m = U_\alpha(r_1,s_1)$$

and consequently prefer contract 1 to contract 0. Secondly, an individual type 0 realizes that welfare under contract 1 is

$$(45) \qquad U_0(r_1,s_1) = m + 2\alpha\sigma^2 \ (s_1/2)^2 = m + \frac{\alpha\sigma^2}{2(1 + 2\alpha\sigma^2)^2}$$

which exceeds the welfare under contract 0,

$$(46) \qquad U_0(r_0,s_0) = m < U_0(r_1,s_1) .$$

In other words: As long as there are type-1 job searchers in the labour market and type-0 agents cannot be excluded from choosing contract 1 (which is designed to type-1 individuals), the reservation utility of type 0 agents is endogenously increased from m to the level (45).

Both comparisons (44), (46) demonstrate that a self-selection device made up of contracts 0,1 will break down.

Fortunately there is a straight-forward revision. Realizing that type 0 agent's reservation level m_0 is now endogenously given through (45),

$$(47) \qquad m_0 = m + \frac{\alpha\sigma^2}{2(1 + 2\alpha\sigma^2)^2}$$

while the reservation of type 1 agents is still at the old level m, the principal could modify contract 0 correspondingly. This modification is called

contract 2 :

share $s_2 := s_0 = 1$

rent $r_2 := r(0, m_0, s_2) = m - \frac{1}{4} + \frac{\alpha\sigma^2}{2(1 + 2\alpha\sigma^2)^2}$.

While the share remains unchanged, the rent refers now to m_0 rather than to m.

Is the set of contracts 1,2 working as a self-selection device? The answer is yes.

To prove this answer one has to consider the choice between contracts 1,2 for each type of individuals. Firstly, contract 2 was constructed in such a way that type-0 agents are indifferent between contracts 1,2. So increase the rent of contract 2 by one dollar or so to induce type-0 agents definitely decide for contract 2. Secondly, type-1 agents still prefer contract 1 when having the choice among contracts 1,2. This follows from

$$(48) \quad U_\alpha(r_2, s_2) = m + \frac{\alpha\sigma^2}{2(1 + 2\alpha\sigma^2)^2} - \frac{\alpha\sigma^2}{2} < m = U_\alpha(r_1, s_1) .$$

One should not forget, however, to ask the principal what she thinks about the self-selection device of contracts 1,2. The final question reads: is the principal really better off under this self-selection device where each agent reveals his type through choice? Comparison is made with respect to the situation before, where everybody just got contract 1.

Let us see the answer: If a type-0 agent decides for contract 2 instead of contract 1, there are two changes of the principal's welfare. Firstly, there is an increase of welfare V according to (43). Secondly, there is the cost $m_0 - m$ associated with the endogenous reservation welfare of type-0 agents. The net effect turns out to be positive.

$$(49) \quad (V_0 - V_\alpha) - (m_0 - m) = \left(\frac{\alpha\sigma^2}{1 + 2\alpha\sigma^2}\right)^2 > 0 .$$

Consequently, the principal prefers to offer the self-selection device. Her incentive to replace the uniform contract 1 by the device of self-selection between contracts 1,2 becomes the greater, the more different job searchers are with respect to their hidden characteristic risk aversion.

Finally, the principal can further increase her net gain (49). She realizes that the costs of the self-selection device, $m_0 - m$, are due to the fact that type-0 agents cannot be excluded from choosing contract 1. So the trick is to modify contract 1 such that the induced increase of type-0 agent's reservation level will not be as much. This is possible, indeed. Since contract 1 maximizes the principal's welfare, a small variation from s_1 to $s_3 := s_1 - \delta$ and from rent r_1 to $r_3 := r(\alpha, m, s_3)$ causes a welfare loss of second order only. On the other hand, the difference $U_0(r_1, s_1) - U_0(r_3, s_3)$ of type-0 agent's welfare is of first order in δ. This means that the principal improves herself when offering self-selection between contracts 2,3 rather than self-selection between contracts 1,2.

As was pointed out by K.J. Arrow (1986) it is typical for problems of hidden characteristics that not all types of searching individuals can find exactly that offer they would get if their characteristics were known by the other market side. This result can be cast in those words: Consider, on one side of the market, "weak" as well as "strong" individuals, characteristics that are hidden to the other market side. Usually, weak individuals need help and must be treated with care. In order to induce strong people to renounce on care and to help themselves, they must get an extra bonus: not for justice, but to set incentives. The size of the bonus depends on the weakness of the weak, or to be more precise, the amount of care devoted to the weak. Sometimes, the weak are not treated with the proper care, just to make the strong peoples' bonus a bit smaller.

References

ALESSI, LOUIS DE: Property Rights, Transaction Costs, and X-Efficiency: An Essay in Economic Theory. American Economic Review 73 (1983) 1, 64-81.

ARROW, KENNETH J.: Agency and the Market. Chapter 23, Handbook of Mathematical Economics, Vol. III, edited by K.J. ARROW and M.D. INTRILLIGATOR, North-Holland 1986, 1183-1200.

ALCHIAN, ARMEN A. and HAROLD DEMSETZ: Production, Information Costs and Economic Organization. American Economic Review 62 (1972) 5, 777-795.

BAMBERG, GÜNTER and KLAUS SPREMANN: Implications of Constant Risk Aversion. Zeitschrift für Operations Research 25 (1981), 205-224.

BAMBERG, GÜNTER and KLAUS SPREMANN: Risikoprämien und Informationswerte bei lokaler Konsistenz. In M.J. BECKMANN et al. (eds.) Mathematische Systeme in der Ökonomie, Athenäum-Verlag 1982, 77-100.

BARNEA, AMIR, ROBERT A. HAUGEN and LEMMA W. SENBET: Agency Problems and Financial Contracting. Prentice-Hall, Englewood Cliffs N.J. 1985.

BLICKLE, MARINA: Information System and the Design of Optimal Contracts. In G. BAMBERG and K. SPREMANN (eds.): Agency Theory, Information and Incentives, Springer Verlag 1987.

FAMA, EUGENE F.: Agency Problems and the Theory of the Firm. Journal of Political Economy 88 (1980) 2, 288-307.

GJESDAL, F.: Information and Incentives: The Agency Information Problem. Review of Economic Studies 49 (1982), 373-390.

GROSSMAN, SANFORD J. and OLIVER D. HART: An Analysis of the Principal-Agent-Problem. Econometrica 51 (1983) 1, 7-45.

HARRIS, MILTON and ARTUR RAVIV: Optimal Incentive Contracts with Imperfect Information. Journal of Economic Theory 20 (1979), 231-259.

HIRSHLEIFER, JACK and JOHN G. RILEY: The Analytics of Uncertainty and Information - An Expository Survey. Journal of Economic Literatur XVII (1979), 1375-1421.

HOLMSTRÖM, BENGT: Moral Hazard and Observability. Bell Journal of Economics 10 (1979), 74-91.

HOLMSTRÖM, BENGT: Moral Hazard in Teams, Bell Journal of Economics 13 (1982), 324-340.

JENSEN MICHAEL C. and WILLIAM H. MECKLING: Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. Journal of Financial Economics 3 (1976), 305-360.

LAZEAR EDWARD P. and SHERWIN ROSEN: Rank-Order Tournaments as Optimum Labour Contracts. Journal of Political Economy 89 (1981), 841-864.

MIRRLEES, JAMES A.: The Theory of Moral Hazard and Unobservable Behaviour - Part I. Nuffield College, Oxford, Mimeo 1975.

MIRRLEES, JAMES A.: The Optimal Structure of Incentives and Authority Within an Organization. Bell Journal of Economics 7 (1976), 105-131.

PRATT, JOHN W.: Risk aversion in the small and in the large. Econometrica 32 (1964), 122-136.

PRATT, JOHN W. and RICHARD J. ZECKHAUSER: Principals and Agents: The Structure of Business. Havard Univ.Press, Boston 1985.

REES, RAY: The Theory of Principal and Agent - Part I. Bulletin of Economic Research 37 (1985) 1, 3-26.

ROSS, STEVEN A.: The Economic Theory of Agency: The Principal's Problem. American Economic Review 63 (1973) 2, 134-139.

SCHMIDT, REINHARD H.: Agency Costs are not a "Flop"! In G. BAMBERG, K. SPREMANN (ed.) Agency Theory, Information and Incentives. Springer Verlag, Berlin 1987.

SCHNEIDER, DIETER: Agency Costs and Transaction Costs: Flops in the Principal-Agent Theory of Financial Markets.In G. BAMBERG and K. SPREMANN (eds.): Agency Theory, Information, and Incentives, Springer Verlag 1987.

SHAVELL, STEVEN: Risk-sharing and Incentives in the Principal-Agent Relationship. Bell Journal of Economics 10 (1979) 55-73.

SHAVELL, STEVEN: On Moral Hazard and Insurance. Quarterly Journal of Economics, 93 (1979) 541-562.

SINGH, NIRVIKAR: Monitoring and Hierarchies: The Marginal Value of Information in a Principal-Agent Model. Journal of Political Economy 93 (1985), 3, 599-609.

SPENCE, MICHAEL and RICHARD J. ZECKHAUSER: Insurance, Information, and Individual Action. American Economic Review 61 (1971) 380-387.

STIGLITZ, JOSEPH E.: Incentives, risk, and information: Notes toward a theory of hierarchy. Bell Journal of Economics 6 (1975), 552-579.

STIGLITZ, JOSEPH E.: Risk Sharing and Incentives in Sharecropping. Review of Economic Studies 61 (1974), 219-256.

STIGLITZ, JOSEPH E. and ANDREW WEISS: Credit Rationing in Markets with Imperfect Information. American Economic Review 71 (1981) 3, 393-410.

SWOBODA, PETER: Kapitalmarkt und Unternehmensfinanzierung - Zur Kapitalstruktur der Unternehmung. Diskussionspapier Graz 1986.

Managerial Contracting with Public and Private Information

Hellmuth Milde

Summary

This paper is concerned with the relationship between a principal and an agent. The principal-agent problem is a special case of the more general moral hazard problem. The basic issue is to design contracts that share risk and simultaneously preserve incentives. The source of moral hazard is the principal's inability to perfectly monitor the agent's actions. These actions together with the state of the world determine the company's cash flow. Moral hazard can be diminished by designing incentive compatible contracts. In order to simplify the analytical solutions we make explicit assumptions regarding the utility function, the density function, and the sharing rule. As a result, the application of mean-variance analysis is possible.

Table of Contents:

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

1. Introduction

The theory of decison making with informational asymmetry examines two closely related problems: adverse selection and moral hazard. Adverse selection is an identification problem. Sometimes commodity markets and financial markets are characterized by asymmetrically distributed quality information. In his founding paper, Akerlof (1970) considers a market in which buyers are unable to ascertain the quality of a used car before the purchase while sellers are aware of the true quality. There is no possibility to transmit reliable quality information from insiders to outsiders. It is easy to demonstrate that this inability might prevent mutually advantageous transactions from taking place. In Akerlof's special case only the lowest quality cars ("lemons") are actually traded. To put it in more general terms, "lemon" markets typically result in serious market failures.

Moral hazard, on the other hand, is an incentive problem. In one of the founding papers, Ross (1973) considers the relationship between a principal and an agent. The basic informational asymmetry is generated by the principal's inability to perfectly (or costlessly) observe and monitor the agent's actions. The importance of moral hazard has been emphazised in the managerial theory of the firm. In modern corporations, there is separation between ownership and control. As owners (or principals) of the firm, stockholders delegate decision-making authority to managers. Managers are the stockholders' agents and their responsibility is to run the company in the stockholders' best interest. However, due to imperfect observability, managers have the opportunity to pursue their own goals, which are in conflict with the goals of stockholders. It is important to note that imperfect observability of managers' actions is the basic source of the moral hazard problem. Even the observable end-of-period cash flow (or wealth) of the corporation does not give a clue about the agent's true contribution. This is because the corporation is operating in an uncertain environment. Consequently, the observable end-of-period value is a mixture of a random term and the manager's input, thus implying the inability to isolate the manager's true contribution. In order to induce the manager to act in the owner's best interest a specific design is required for the manager's compensation schedule. The purpose of this design is to achieve incentive compatibility. A contractual arrangement, which is incentive compatible, guarantees the highest possible manager input and the best possible result for the principal by acting as a self-enforcing mechanism.

This paper is about incentive compatibility of managerial compensation schemes. The next section outlines the major assumptions of the basic model. In section 3 we focus on a model with public information. In the public information case the contract, which is endogenously derived, is called a first best solution. We give a detailed discussion of this case in order to have a benchmark to evaluate models with differing informational assumptions. In section 4 we assume that information is private. We shall discuss the structuring of a contract which is denoted as a second best solution. We also compare both contracts and demonstrate the effects of asymmetric information on managerial contracting.

Basically we shall follow the framework of Ross (1973), Holmstrom (1979), Stiglitz (1983) and Chapter III of the 1985 book by Barnea/Haugen/Senbet (BHS). While Ross, Holmstrom, and Stiglitz do not make special assumptions regarding the underlying functions, it is well known that analytical solutions are simplified under explicit assumptions regarding the utility function and the density function. In this paper we follow BHS's advice (p.29) and use a CARA utility function (Constant Absolute Risk Aversion) in combination with normally distributed cash flows and linear sharing rules. This allows the application of the mean-variance approach, and guarantees explicit solutions for some of the contracts discussed in this paper.

2. Assumptions

2.1 Risk Preference Function

When we analyse decision making under uncertainty we think in terms of expected utility. In order to specify the underlying utility function we usually assume that the average decision maker is risk averse. Risk aversion implies a concave utility function. In most of the literature dealing with portfolio problems or agency problems the only assumption made is that utility is an increasing and concave function of the end-of-period wealth. However, imposing more structure on the utility function provides the possibility to derive explicit solutions. Following BHS we assume that both principal and agent make decisions according to a CARA utility function. In this case, the manager's utility U is given by

$$U(R) = 1 - \exp(-A R) \quad (1)$$

and the principal's utility V is given by

$$V(Q) = 1 - \exp(-B Q), \quad (2)$$

where

$\tilde{R}$ = manager's stochastic(end-of-period)income and R is the realization,

$\tilde{Q}$ = principal's stochastic(end-of-period)income and Q is the realization,

A = manager's coefficient of absolute risk aversion ($A > 0$),

B = principal's coefficient of absolute risk aversion ($B > 0$).

One reason that CARA utility functions have proved popular in practical examples is the absence of wealth (income) effects. Using the definition of the Arrow-Pratt-measure of absolute risk aversion it is easy to show that A and B are the constant risk aversion parameters:

$$- U_{RR}(R)/U_R(R) = A,$$

$$- V_{QQ}(Q)/V_Q(Q) = B.$$

2.2 Sharing Rules

Next we explain the definitions of the manager's income $\tilde{R}$ and the principal's income $\tilde{Q}$. Consider a firm with separation of ownership and control operating in a single period framework. The end-of-period cash flow (or wealth) is $\tilde{X}$, where X, the realization of the random variable $\tilde{X}$, is unknown to both principal and agent. The analysis is presented in three steps: Firstly, managerial effort is assumed to have no impact on the realization of $\tilde{X}$, thus making the distribution of $\tilde{X}$ an exogenously given function. Moreover, we assume that the distribution is symmetrically known to all parties. Secondly, following the standard agency literature the expected value of $\tilde{X}$ is affected by the agent's effort. The effort is assumed to be observable by the principal. Thirdly, as before the distribution of $\tilde{X}$ is governed by the agent's input, however, the true input is not observable. Details regarding the density function of $\tilde{X}$ are discussed in the next subsection.

Given that both principal and agent are risk averse, it is obvious that the two parties cannot both achieve certainty situations. Consequently, some proportion of the social risk will be borne by each party. In this paper we shall discuss possible arrangements of risk sharing. As indicated, effort is introduced step by step. The impact of different assumptions regarding effort on the optimal solution of risk sharing is analysed.

As to the sharing rule we assume the simplest possible case. Only linear (or proportionate) sharing rules are considered. If the sharing rule is linear, the manager will get a constant fraction α of the company's cash flow. The residual fraction $(1-\alpha)$ will be left for the principal. In general, a linear sharing rule is not an optimal strategy in the Pareto sence. However, Pareto optimality can be achieved if side payments from one party to the other are taken into account. In our model the side payment is the manager's fixed income component F. Both α and F are the owners' decision variables. Any scheme of the manager's compensation is completely described by a pair $\{\alpha, F\}$.

Given alternative pairs of α and F the manager's stochastic income is given by

$$\tilde{R} = \alpha \tilde{X} + F, \tag{3}$$

and the principal's income is given by

$$\tilde{Q} = (1-\alpha) \tilde{X} - F, \tag{4}$$

where $0 \leq \alpha \leq 1$ and $F \gtrless 0$.

2.3 Density Function

In this paper we assume that the firm's cash flow $\tilde{X}$ is normally distributed. Thus, the density function is given by

$$N(\tilde{X}; \mu, \sigma^2),$$

where both moments $E[\tilde{X}] = \mu$ and $Var[\tilde{X}] = \sigma^2$ are exogenously given in the basic risk sharing model (subsection 3.1). Later (subsection 3.2 and section 4) we shall drop this assumption and introduce a functional relationship between the expected value of $\tilde{X}$ and the managers' input. The properties of the "production" function are discussed in the next subsection.

As is well known, the combination of a CARA utility function with a normally distributed random variable results in a mean-variance model. The popularity of the mean-variance approach is due to the property that risk is fully described by the variance of the random variable. According to our assumption, $\tilde{X}$ is normally distributed. However, X is not the relevant argument in the utility functions. As argued in (1) and (2) the utilities are functions of R and Q, respectively. On the other hand, it is well known that any linear combination of a normally distributed random variable is also normal. Thus, the linear sharing rule discussed in (3) and (4) guarantees that the normal distribution of $\tilde{X}$ is preserved for the cash flows $\tilde{R}$ as well as $\tilde{Q}$. Without the introduction of the linear rule $\tilde{R}$ and $\tilde{Q}$ would not be normally distributed in general.

Given the normal distribution of $\tilde{R}$ and $\tilde{Q}$, the combination with (1) and (2) results in the following proposition: The maximization of expected utilities $E[U(\tilde{R})]$ and $E[V(\tilde{Q})]$ is identical to the maximization of the associated certainty equivalents G and H (for details see Parkin/Gray/Barrett (1970) or Bamberg (1986)), where

$$G = E[\tilde{R}] - A\ Var[\tilde{R}]/2, \tag{5}$$

$$H = E[\tilde{Q}] - B\ Var[\tilde{Q}]/2. \tag{6}$$

From (3) and (4) we derive the following moments:

$$E[\tilde{R}] = \alpha\mu + F, \quad Var[\tilde{R}] = \alpha^2\ \sigma^2, \tag{7}$$

$$E[\tilde{Q}] = (1-\alpha)\mu - F, \quad Var[\tilde{Q}] = (1-\alpha)^2\sigma^2. \tag{8}$$

2.4 Production Function

In our earlier discussion we argued that the expected cash flow will be a function of the manager's effort. Following the standard agency model we assume that effort affects the random variable according to the criterion of first-order stochastic dominance. As it is well known one random variable is dominated by a second random variable in the sense of first-order stochastic dominance if the second random variable differs from the first random variable by a positive additive amount; see Kadar/Russel (1971) for details. Thus, introducing a standard production function we assume that more effort e generates a "better" random variable where "better" means dominance in the first-order sense.

The production function is assumed to have well-known properties, i.e. the expected cash flow is increasing with increasing effort but shows diminishing marginal returns:

$$\mu = \mu(e), \text{ with} \tag{9}$$

$$\mu_e > 0, \ \mu_{ee} < 0.$$

Having specified the production function we proceed by decomposing the random cash flow $\tilde{X}$. In order to simplify the analysis we assume that μ enters the definition of the new (standardized) random variables $\tilde{Y}$ additively:

$$\tilde{Y} = \tilde{X} - \mu \quad . \tag{10}$$

As a result, $\tilde{Y}$ is distributed normally with $N(\tilde{Y}; 0, \sigma^2)$, i.e. the expected value of $\tilde{Y}$ is zero and the variance of $\tilde{Y}$ is the same as that of $\tilde{X}$.

3. A Model with Public Information

3.1 The Nature of Risk Sharing

In this section we discuss a model with public information. The first subsection ignores managerial effort completely. In the second subsection effort is assumed to have an impact on the expected cash flow. Moreover, the effort is observable by assumption. Analyzing the risk sharing problem in isolation, a contract is fully described by alternative combinations of the decision variables α and F. Note that A, B, μ and σ are exogenously given data. The objective functions of the principal and the agent, respectively, are derived from (5) to (8) and (10). The agent's function is given by

$$G = \alpha\mu + F - A \ \alpha^2 \ \sigma^2/2, \tag{11}$$

with

$$G_\alpha = \mu - A \sigma^2 \alpha \gtrless 0, \quad G_{\alpha\alpha} = - A \sigma^2 < 0,$$

$$G_F = + 1, \quad G_{FF} = G_F = 0;$$

the principal's function is given by

$$H = (1 - \sigma)\ \mu - F - B\ (1 - \alpha)^2\ \ \sigma^2/2, \tag{12}$$

with

$$H_\alpha = \ -\mu + B\ \sigma^2(1 - \alpha) \gtreqless 0, \quad H_{\alpha\alpha} = -\ B\ \sigma^2 < 0,$$

$$H_F = -\ 1, \quad H_{FF} = H_{F\alpha} = 0.$$

We proceed by analyzing the preference maps of a representative principal and an agent. Starting with the agent's situation we can derive and show the following proposition: In an α - F space the agent's indifference curve is convex with a unique turning point (minimum) at $G_\alpha = 0$. To see this, note the following properties of the indifference (iso-G) curve:

$$\left.\frac{dF}{d\alpha}\right|_G = -\ G_\alpha/G_F = -\ \mu\ +\ A\,\alpha\,\sigma^2 \gtreqless 0, \tag{13a}$$

$$\left.\frac{d^2F}{d\alpha^2}\right|_G = -G_{\alpha\alpha}\ /G_F = +\ A\ \sigma^2 > 0. \tag{13b}$$

As to the indifference curve of the principal we show the following proposition: In an α- F space the principal's indifference curve is concave with a unique turning point (maximum) at $H_\alpha = 0$. To see this, note the following properties of the indifference (iso-H) curve:

$$\left.\frac{dF}{d\alpha}\right|_H = -\ H_\alpha\ /H_F = -\ \mu + B\,\sigma^2(1 - \alpha) \gtreqless\ 0, \tag{14a}$$

$$\left.\frac{d^2F}{d\alpha^2}\right|_H = -\ H_{\alpha\alpha}\ /H_F = -\ B\ \sigma^2 < 0. \tag{14b}$$

The terms (13a) and (14b) are marginal rates of substitution between α and F. The marginal rate of substitution is the maximum or minimum amount of F required to compensate for a marginal change in α in order to keep G or H at an exogenously given level. Any α- F combination is characterized by a specific level of H and G. All α - F combinations with non-negative H and G levels are acceptable for both parties. Thus, the set of feasible α - F combinations is bounded by the zero-H- and zero-G-indifference curve.

Next we ask which of the possible α - F combinations are optimal in the Pareto sense. As is known from basic propositions of welfare economics a Pareto efficient solution is characterized by identical marginal rates of substitution of the parties involved. Thus, we are looking for all α- F combinations at which the indifference curves of principals and agents are tangential (contract curve). The condition satisfying Pareto efficiency is given by

$$\left.\frac{dF}{d\alpha}\right|_G = \left.\frac{dF}{d\alpha}\right|_H \tag{15a}$$

or

$$-\mu + A\,\sigma^2\,\alpha^* = -\mu + B\,\sigma^2(1 - \alpha^*). \tag{15b}$$

From (15) we derive the agent's optimal fraction α^* and the principal's optimal fraction $(1 - \alpha^*)$:

$$\alpha^* = B/(A + B), \tag{16a}$$

$$1 - \alpha^* = A/(A + B). \tag{16b}$$

According to (16a) and (16b) the fractions depend on the exogenously given and publicly known coefficients of absolute risk aversion. If both partners have identical coefficients, i.e. A = B, we find $\alpha^* = (1 - \alpha^*) = 1/2$. On the other hand, if the manager is risk neutral, A = 0, we derive $\alpha^* = 1$. This means that the manager performs the role of an insurance company which takes over all risk. In this situation the owner sells the risky firm to the managers. More generally, we derive the following comparative static results: An increase in B or a decrease in A will increase the optimal fraction α^*. Indicating elasticities by ε we obtain

$$\varepsilon(\alpha^*, A) = -A/(A + B) < 0, \tag{17a}$$

$$\varepsilon(\alpha^*, B) = +A/(A + B) > 0. \tag{17b}$$

Note that a simultaneous change of both A and B by the same rate will not change α^*:

$$\varepsilon(\alpha^*, A) + \varepsilon(\alpha^*, B) = 0.$$

The reader will no doubt have noticed that F does not occur in (15) and (16). This result is intuitively clear if we recall the definition of Pareto optimality.

A Pareto optimum is a situation with a maximum of social welfare (or total gain, G + H). From (11) and (12) we get

$$G + H = \mu - (A\alpha^2 + B(1-\alpha)^2)\,\sigma^2/2, \tag{18}$$

which shows that the term F drops out. One more implication can be obtained from (18). The substitution of (16a) and (16b) into (18) results in the highest possible social gain

$$G^* + H^* = \mu - AB\,\sigma^2/2(A + B) > 0. \tag{19}$$

In (19) we assume that the highest possible social gain is a positive number. This implies

$$\mu > AB\,\sigma^2/2(A + B), \tag{20}$$

which imposes some constraints on the choice of the parameters of the model.

Not surprisingly, we can obtain the fractions α^* and $(1 - \alpha^*)$ also from maximizing the social gain function (18) with respect to α. The first and second order conditions are given by

$$(G+H)_\alpha = -(A\,\alpha^* - B(1-\alpha^*))\,\sigma^2 = 0, \tag{21a}$$

$$(G+H)_{\alpha\alpha} = -(A + B)\sigma^2 < 0. \tag{21b}$$

The first order condition (21a) confirms the results in (16a) and (16b). The second order condition (21b) is automatically satisfied.

Given α^*, the optimal amount F^* is obtained from either (11) or (12) depending on the assumptions regarding the degree of competition among the managers. Following the agency literature we assume that the manager market is perfectly competitive which implies G = 0. The market entry and exit of managers drive the long run income and the associated certainty equivalent down to zero. The optimal amount of F turns out to be negative which means that the side payment goes from managers to owners. From (11), (16a), (20) along with G = 0 we obtain

$$F^*_N = -\frac{B}{A+B}\left(\mu - \frac{AB\sigma^2}{2(A+B)}\right) = -\alpha^*(G^* + H^*) < 0 .$$

On the other hand, if managers are in a strong position the owner's income is driven down to zero which implies H = 0. In this situation the managers receive the highest possible amount of fixed income

$$F^*_M = -\frac{A}{A+B}\left(\mu - \frac{AB\sigma^2}{2(A+B)}\right) = (1-\alpha^*)(G^* + H^*) > 0 \ .$$

In Figure 1, the possible cases are depicted. Moreover, the reader may wish to refresh his memory by comparing the propositions derived in this subsection with the features in Figure 1.

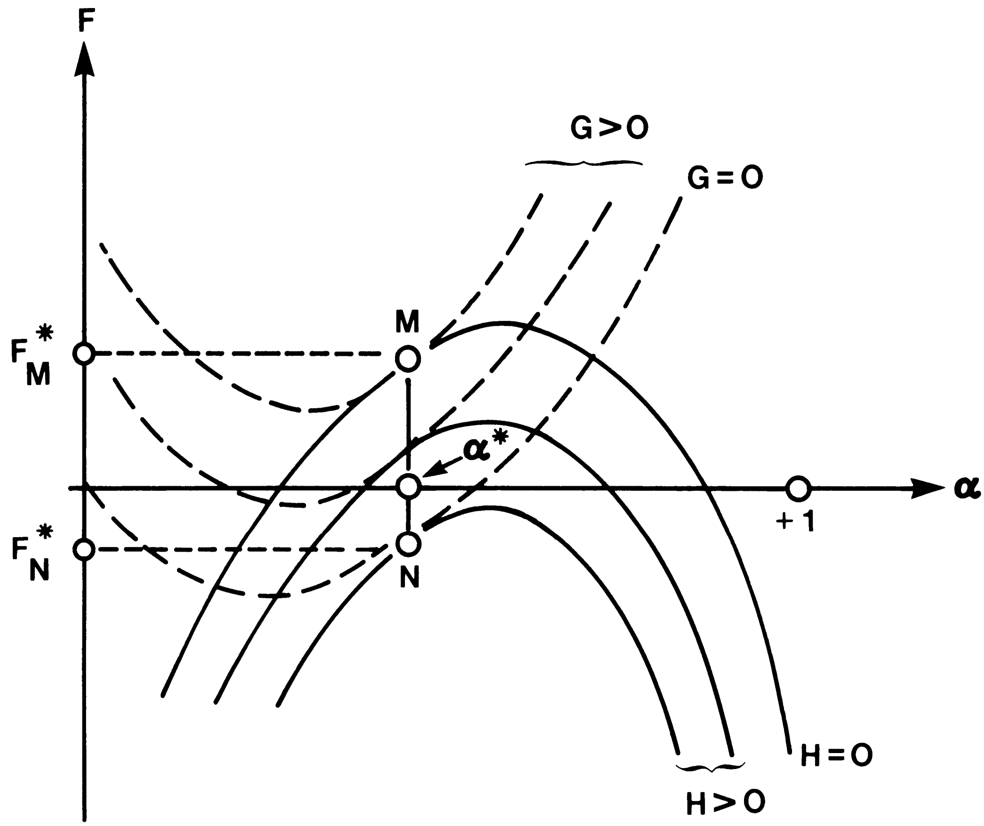


FIGURE 1: Preference Maps of Principals and Agents

In order to demonstrate that the solutions (16) and (21) are based on a common approach we briefly discuss the more general Lagrangean method. According to an alternative definition a Pareto optimum is a situation with the property that one individual cannot be made better off without making someone else worse off. We continue to consider the choice in terms of certainty equivalents. The principal's choice is to maximize H, but to maximize subject to a constraint $G = \bar{G}$. The decision variables are α and F. The Lagrangean expression L is given by

$$L = H(\alpha, F) + \lambda (G(\alpha, F) - \bar{G}), \qquad (22)$$

where λ is the Lagrangean multiplier which is an additional endogenously determined variable. The first order conditions are given by

$$L_\alpha = H_\alpha + \lambda G_\alpha = 0, \qquad (23a)$$

$$L_F = H_F + \lambda G_F = 0, \qquad (23b)$$

$$L_\lambda = G - \bar{G} = 0. \qquad (23c)$$

Eliminating λ, we find from (23a) and (23b)

$$-\frac{G_\alpha}{G_F} = -\frac{H_\alpha}{H_F},$$

which is exactly (15a) using (13a), (14a). Taking the results G_F and H_F from (11) and (12) explicitly into account we obtain the result already known from (15b),

$$\lambda = 1 \quad \text{and} \quad -G_\alpha = H_\alpha .$$

Using $\lambda = 1$ in (22) we obtain

$$L = H(\alpha, F) + G(\alpha, F) - \bar{G}.$$

which is identical to (18) up to the constant term $\bar{G}$. The maximization of L gives the same results as the maximization of G + H, see (21).

3.2 Risk Sharing and Observable Effort

This subsection assumes that managerial effort e has a measurable impact on μ. Moreover, suppose that effort is observed by all parties. In this case, a managerial contract will relate the manager's compensation to his actual (and observable) effort. The effort is determined by the principal. In order to guarantee correct managerial incentives, the contract might include penalties if the actual effort is below the agreed-upon optimal effort.

The principal's objective function is a combination of (9) and (12):

$$H = (1-\alpha)\ \mu(e) - F - B(1-\alpha)^2\ \sigma^2/2. \tag{24}$$

Note that μ is no longer an exogenously given parameter. According to (9) the input of managerial effort e determines μ.

The manager's objective function requires additional modification. It is important to take the manager's opportunity cost explicitly into account. If the manager is not working in the risky company under consideration, he has the alternative to invest his effort in a riskfree job earning the riskfree (and exogenously given) wage rate w. Thus, his opportunity cost is we. The manager's stochastic profit $\tilde{P}$ is given by

$$\tilde{P} = \alpha \tilde{X} + F - w\,e, \quad \text{with}$$

$$E[\tilde{P}] = \alpha\mu + F - w\,e\ ,\ Var[\tilde{P}] = \alpha^2\ \sigma^2.$$

The certainty equivalent of the manager's expected utility of profit, denoted by S, is given by

$$S = E[\tilde{P}] - A\ Var[\tilde{P}]/2 = \alpha\mu + F - we - A\ \alpha^2\ \sigma^2/2.$$

Alternatively, we can derive S from (5) or (11). Thus, we obtain

$$S = G - w\,e = \alpha\mu + F - A\ \alpha^2\ \sigma^2/2 - w\,e.$$

Taking (9) into account the manager's objective function is given by

$$S = \alpha\mu(e) + F - we - A\ \alpha^2\ \sigma^2/2. \tag{25}$$

In order to find the Pareto optimal contract we solve the principal's optimization problem. The three endogenous variables are α, F and e. Formally, the problem is given by

$$\max_{\alpha,\ F,\ e}\ H = (1-\alpha)\ \mu\ (e) - F - B(1-\alpha)^2\ \ \sigma^2/2 \tag{26a}$$

$$\text{subject to } S = \alpha\mu(e) + F - we - A\ \alpha^2\ \sigma^2/2 \geq\ 0. \tag{26b}$$

The constraint guarantees the manager a non-negative certainty equivalent. If S is negative the contract is not acceptable for the manager. As a result, he will choose the alternative employment with the riskfree income w e. We proceed by following the standard agency literature and assume that the contraint (26b) is binding: S = 0. This means competition in the manager market drives long run profits down to zero.

As was pointed out, the general method to solve a constrained optimization problem is the Lagrangean approach. However, to simplify the procedure we shall employ the "substitution" method. We solve the constraint (26b) for F and substitute the result into the objective function (26a). The modified objective function is given by

$$H = \mu(e) - we - (A\alpha^2 + B(1-\alpha)^2)\,\sigma^2/2. \tag{27}$$

Note that F is eliminated as a result of the substitution. The remaining endogenous variables are e and α. The first-order conditions are given by

$$H_e = \mu_e(e^*) - w = 0, \tag{28}$$

$$H_\alpha = -(A\,\alpha^* - B(1-\alpha^*))\,\sigma^2 = 0. \tag{29}$$

Second-order conditions are satisfied:

$$H_{ee} = \mu_{ee} < 0, \tag{30a}$$

$$H_{ee}\,H_{\alpha\alpha} - H_{e\alpha}^2 = - \mu_{ee}\,(A + B)\,\sigma^2 > 0. \tag{30b}$$

The results are straightforward. According to (28) the manager's effort has an optimal level e^* if the marginal expected return μ_e is equal to the marginal cost w. This condition is known from the theory dealing with firm behavior under certainty. The second optimality condition (29) which determines α^* is known from (21a) or (15b). Thus, the results (16) and (17) hold as well.

The optimal amount F^* is calculated from (26b). Although we continue to assume perfectly competitive manager markets as in subsection 3.1, we obtain a result different from F^*_N. The reason is that opportunity cost is now explicitly taken into account. The important insight derived in this subsection is given by the following proposition: Observable effort does not change the optimal fraction α^* of the risk sharing contract.

4. A Model with Private Information

4.1 The Manager's Decision Problem

We now depart from the assumption of observable effort and assume instead that the principal is no longer able to observe the manager's effort. Given the inability to monitor or observe the agent's actions there is no way to assign responsibilities. A low end-of-period cash flow of the company is either the result of insufficient manager input or of a bad state of the world. A distinction is impossible. Furthermore, it no longer makes sense to relate the manager's compensation to his actual but unobservable effort. In contrast to the setting discussed in the last subsection, effort is no longer a variable determined by the principal. Any contract is fully described by $\{\alpha, F\}$. Thus, we return to contracts discussed in subsection 3.1 although now effort does influence the expected value of $\tilde{X}$.

In the present context the manager (and not the principal) decides what is the optimal amount of effort. The manager-determined optimal effort, denoted e^o, is the result of the manager's maximization process. The manager's objective function is given by (25). For the first and second order conditions we obtain

$$S_e = \alpha \mu_e (e^o) - w = 0, \tag{31a}$$

$$S_{ee} = \alpha \mu_{ee} (e^o) < 0. \tag{31b}$$

According to the first order condition (31a) the optimal effort depends on both w and α. The wage rate is given by the market. On the other hand, the fraction α is a choice variable of the principal. Changes in α will result in systematic reactions of the optimal effort e^o. The relationship between α and e^o is described by the reaction function

$$e^o = e^o(\alpha). \tag{32a}$$

The property of (32a) is derived as the result of a simple comparative static exercise. From (31a) and (31b) we obtain

$$e^o_\alpha = - \mu_e (e^o) / \alpha \mu_{ee} (e^o) > 0. \tag{32b}$$

The result in (32b) is intuitively clear. An increase in the fraction α will increase the manager's effort. The reaction function (32a) and the property (32b) are assumed to be publicly known to all parties.

4.2 The Principal's Decision Problem

As was pointed out, effort is no longer a decision variable of the principal.

However, knowing (32a) and (32b) a "sophisticated" principal will not ignore this information when it comes to setting the optimal values of α and F. As demonstrated in subsection 4.1 the choice of α implies a specific level of e^o, which, according to (9), determines the expected value μ of the density function. Consequently, an increase of α generates a "better" density function (where "better" is defined in the sense of first order stochastic dominance), thus shifting the density function to the right:

$$\mu_\alpha = \mu_e \, e^o_\alpha > 0. \tag{33}$$

Without going into details we infer from (33) that the fraction α now performs two different economic functions. In contrast to section 3, α is not just an instrument to allocate a given level of social risk among the two parties. In addition, α now generates incentives to modify the riskiness of the density function. The modification of social risk and the incentive to do so are the essential features of moral hazard.

Next we ask what are the terms of an optimal compensation schedule. As pointed out a contract is fully characterized by the vector $\{\alpha , F\}$. The optimal solutions of α and F are derived from the principal's optimization behavior. The principal's objective function is given by (24) or (26a). However, now the optimization procedure is subject to two constraints. As already demonstrated in (26b), the first constraint guarantees the managers a minimum certainty equivalent. In addition, the second constraint reflects the manager-determined reaction function (32a). Formally, the optimization problem is given by

$$\max_{\alpha , F} \quad H = (1 - \alpha) \, \mu(e^o) - F - B (1 - \alpha)^2 \sigma^2/2, \tag{34a}$$

$$\text{subject to} \quad S = \alpha \, \mu(e^o) + F - w \, e^o - A \, \alpha^2 \sigma^2/2 = 0 \tag{34b}$$

$$\text{and} \quad e^o - e^o(\alpha) = 0. \tag{34c}$$

As before we assume that (34b) is a binding constraint, thus we obtain S = 0. We proceed by solving the constraint (34b) for F and substitute the term into the objective function (34a). The second constraint is simply substituted in (34a). Note that after eliminating F the only remaining decision variable of the principal is α . The optimal fraction is denoted by α^o. The principal's modified objective function is given by

$$H = \mu(e^o(\alpha)) - w\, e^o(\alpha) - (A\,\alpha^2 + B(1-\alpha)^2)\,\sigma^2/2. \tag{35}$$

The first and second order conditions are given by

$$H_\alpha = (\mu_e(e^o(\alpha^o)) - w)\, e^o_\alpha(\alpha^o) - (A\alpha^o - B(1-\alpha^o))\,\sigma^2 = 0, \tag{36}$$

$$H_{\alpha\alpha} = (\mu_e - w)\, e^o_{\alpha\alpha} + (e^o_\alpha)^2\, \mu_{ee} - (A+B)\,\sigma^2 < 0. \tag{37}$$

The first order condition (36) determines the optimal fraction α^o. By imposing some constraints on the parameters, the second order condition (37) is satisfied.

In order to gain more insights in the optimal solution, we simplify the first order condition (36) by decomposing the marginal expected return

$$\mu_e = (1-\alpha^o)\,\mu_e + \alpha^o\,\mu_e. \tag{38}$$

Substituting (38) and the manager's first order condition (31a) into the principal's optimality condition (36) we obtain

$$H_\alpha = (1-\alpha^o)\,\mu_e\,(e^o(\alpha^o))\, e^o_\alpha(\alpha^o) - (A\,\alpha^o - B(1-\alpha^o))\,\sigma^2 = 0. \tag{39}$$

From (39) we derive an implicit solution for the optimal fraction α^o:

$$\alpha^o = (B + \mu_e\, e^o_\alpha / \sigma^2) / (A + B + \mu_e\, e^o_\alpha / \sigma^2). \tag{40}$$

Comparing (40) with (16a) we obtain the following results: The numerator and the denominator of α^o are larger by the positive amount $\mu_e\, e^o_\alpha / \sigma^2$ (the term is positive because of (9) and (32b)). Alternatively, we can say: compared with (16a) the term B in (40) is larger by the amount $\mu_e\, e^o_\alpha/\sigma^2$. As argued in (17b) an increase in B will result in an increase in α. Consequently, we obtain $\alpha^o > \alpha^*$, i.e. the manager's fraction is larger in case of unobservable effort than in case of observable effort. We continue to assume $A > 0$ which results in an exact specification of the relevant α^o - interval:

$$\alpha^* < \alpha^o < 1.$$

Knowing the interval of possible α^o - fractions, we are in a position to compare e^* with e^o. The effort e^* in the public information case is known from (28), e^o in the private information case is known from (31a). As $\alpha = \alpha^o < 1$ we infer that

$e^o < e^*$. The manager's effort in the private information case is smaller than that in the public information case, see Figure 2.

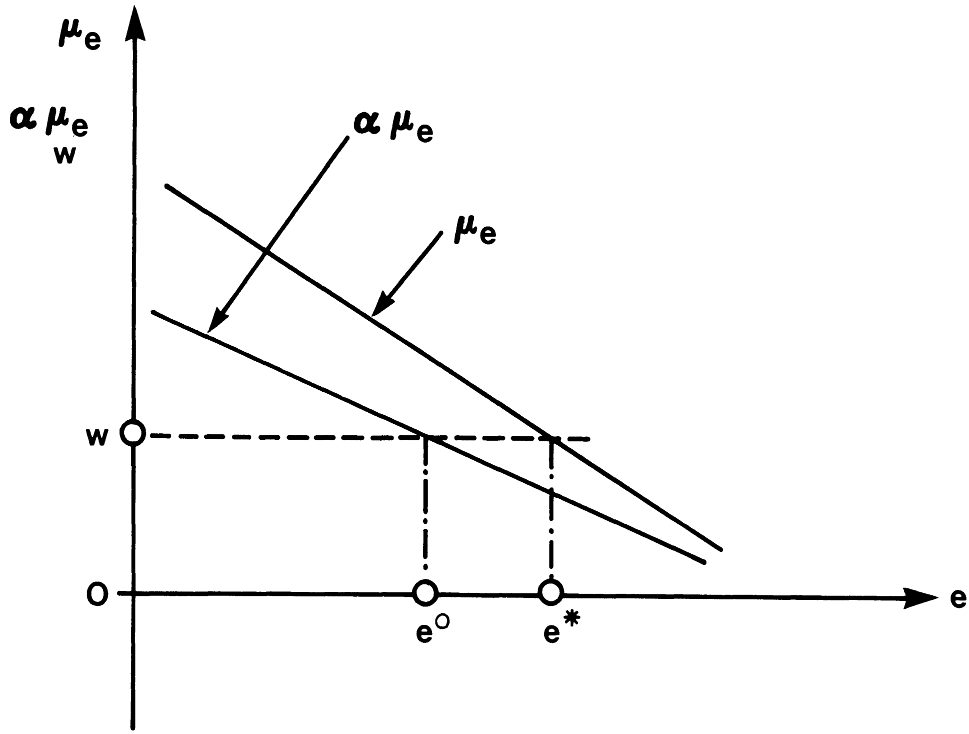


FIGURE 2: Managerial Effort with Public and Private Information

Taking (9) into account we derive $\mu^o < \mu^*$, i.e. the expected value of $\tilde{X}$ is smaller in the private information case than in the public information case. The misallocation of effort and the reduction in the expected cash flow are the results of imperfect observability.

Knowing α^o and e^o the fixed income component F^o under private information can be obtained from (34b). However, without imposing more structure on the production function (9) we do not know whether F^o is larger or smaller than F^*.

In the final subsection we present some more intuition to compare the two contracts with public and private information. As pointed out in the Introduction, the public information contract is denoted as first best solution; the private information contract is denoted as second best solution.

4.3 Comparing the Contracts

In this subsection we re-examine the first best and second best solutions by introducing the concept of cost of social risk. Basically, the cost term occurs in the objective functions (18), (27), and (35). The function of social cost, denoted $C(\alpha)$, is given by

$$C(\alpha) = (A\,\alpha^2 + B(1-\alpha)^2)\,\sigma^2/2, \tag{41}$$

with

$$C_\alpha\,(\alpha) = (A\,\alpha - B(1-\alpha))\,\sigma^2 \gtreqless 0, \tag{42a}$$

$$C_{\alpha\alpha}\,(\alpha) = (A + B)\,\sigma^2 > 0. \tag{42b}$$

Thus, the cost function is convex with a unique turning point (minimum) at $C_\alpha = 0$. Using (41) the objective function (27) in the public information case is given by

$$H = \mu(e) - we - C(\alpha).$$

As a result, the first best contract is characterized by a clear assignment of the decision variables e and to their respective targets. The scale problem is solved by setting e^*, with $\mu_e(e^*) = w$ as the optimality condition. The problem of risk allocation is solved by choosing α^*, so that the cost function is minimized: $H_\alpha\,(\alpha^*) = -\,C_\alpha(\alpha^*) = 0$.

In the private information case the objective function (35) can be simplified to

$$H = \mu(e^O(\alpha)) - w\,e^o(\alpha) - C(\alpha).$$

Note that the second best contract is characterized by just one decision variable: α is used to solve both the scale problem and the risk allocation problem. A change in α has two types of effects. Firstly, as before the allocation of risk is changed (allocation effect). Secondly, as argued in (32), managers react to changes in α by changing their effort (scale effect). As a consequence, there is a change in the expected cash flow.

To be more specific, suppose there is an increase in α. The increase in the expected cash flow is given by $\mu_e(e^o)\,e^o_\alpha\,(\alpha) > 0$. The manager's fraction of the increased cash flow is given by $\alpha\mu_e(e^o)\,e^o_\alpha\,(\alpha)$ which is matched in the optimal situation by a simultaneous increase in the manager's opportunity cost, $w\,e^o_\alpha\,(\alpha)$. Basically, the result is known from (31a):

$$\alpha^{o} \mu_{e}(e^{o})\ e_{\alpha}^{o}\ (\alpha^{o}) = w\ e_{\alpha}^{o}(\alpha^{o}). \tag{43}$$

The remaining fraction $(1 - \alpha)\ \mu_e(e^o)\ e_\alpha^o(\alpha) > 0$ is the principal's increase in the expected cash flow. According to the optimum condition (39) the principal's marginal return is matched by an equal increase in the social cost of risk, $C_\alpha(\alpha)$:

$$(1 - \alpha^o)\ \mu_e(e^o)\ e_\alpha^o\ (\alpha^o) = C_\alpha\ (\alpha^o). \tag{44}$$

As pointed out, the LHS of (44) is a positive number. Consquently, $C_\alpha(\alpha^o)$ on the RHS must be positive as well. Because of the convexity of $C(\alpha)$, the marginal cost is positive if and only if $\alpha^o > \alpha^*$ is satisfied. The second best fraction α^o (derived under private information) lies in the solid interval of Figure 3.

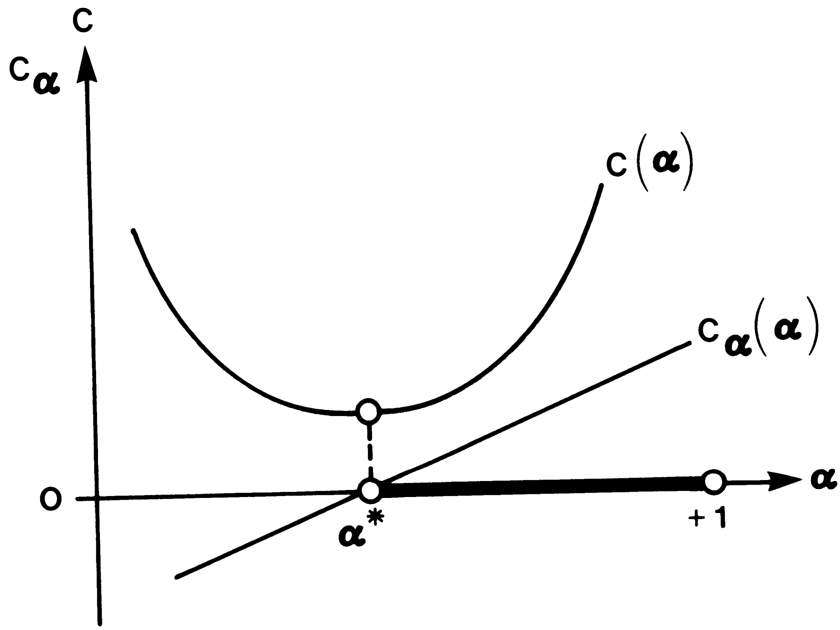


FIGURE 3: Cost of Social Risk as a Function of Risk Sharing

Any increase in α beyond the first best fraction α^* (derived under public information) has two counteracting effects. According to the scale effect, there is a resulting increase in managerial effort (and expected cash flow), thus diminishing the gap between the first best effort e^* and the second best effort e^o. The reduction of misallocation in effort is clearly a social gain. This social gain is the gain enjoyed by the principal. On the other hand, the increase in α beyond α^* changes the allocation of risk for the worse. This deterioration of risk allocation is clearly a social loss. The condition characterizing the second best risk sharing α^o says that the marginal social gain must be equal to the marginal social loss. The two opposite effects balance each other at the margin.

References

Akerlof, G. (1970): The Market For Lemons: Quality Uncertainty and the Market Mechanism, Quarterly Journal of Economics 90, 488-500.

Bamberg, G. (1986): The Hybrid Model and Related Approaches to Capital Market Equilibria, in: Bamberg, G.; Spremann, K. (Eds.): Capital Market Equilibria, Berlin-Heidelberg-New York-Tokyo, 7-54.

Barnea, A.; Haugen, R.A.; Senbet, L.W. (1985): Agency Problems and Financial Contracting, Englewood Cliffs, N.J.

Hadar, J.; Russell, W.J. (1971): Stochastic Dominance and Diversification, Journal of Economic Theory 3, 280-303.

Holmstrom, B. (1979): Moral Hazard and Observability, Bell Journal of Economics 10, 74-91.

Parkin, J.M.; Gray, M.R.; Barrett, R.J. (1970): The Portfolio Behaviour of Commercial Banks, in: Hilton, R; Heathfield, D.J. (Eds.): The Econometric Study of the United Kingdom, London-Basingstoke, 221-239.

Ross, S.A. (1973): The Economic Theory of Agency: The Principal's Problem, American Economic Review (Papers and Proceedings) 63, 134-139.

Stiglitz, J.E. (1983): Risk, Incentives, and Insurance: The Pure Theory of Moral Hazard, Geneva Papers on Risk and Insurance 8, 4-33.

Risk Sharing and Subcontracting

Günter Bamberg

Summary: Linear risk sharing provisions between companies and supply industry are considered. The provisions are characterized by target profit, target cost, and a sharing rate. The problem to assess these parameters appropriately is dealt with in a normative model. The model is parsimoniously parameterized and allows explicit solutions with respect to all contractual parameters. The simplicity of the model makes it possible to incorporate additional aspects such as diversification, heterogeneous expectations or cost monitoring expenditure.

Table of Contents

1. Introduction

Up to now the literature about risk sharing provisions in contracting has been dominated by defense contracting problems. The relevant literature includes Blanning/Kleindorfer/Sankar (1982), Cross (1968), Cummins (1977), Fisher (1969), Fox (1974), Moore (1967), Peck/Scherer (1962), Scherer (1964a,1964b), Tashjian (1974), Williamson (1967). The bias towards military procurement problems has at least two reasons. Firstly, defense contracting consumes a considerable part of the defense budget in most countries. According to Cummins (1977), contracting for national defense has accounted for over half of the total annual US defense budget since the early 1960s. Secondly, in

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

1962, the US Defense Department revised the Armed Services Procurement Regulations to stimulate the use of incentive contracts instead of cost-plus-fixed-fee contracts.

Obviously, military procurement is an important area for the application of incentive contracts. But nonmilitary areas are or could be of equal importance. Kawasaki/McMillan (1985) give figures about the Japanese subcontracting structure. In the Japanese automobile industry, for instance, an average of 75 percent of a car's value is provided by outside suppliers. Moreover, the amount of subcontracting in Japanese manufacturing industry is increasing over time. Though Western Europe and the US show a less pronounced (nonmilitary) subcontracting structure, there seems to be some need for easy applicable incentive contracts.

The incentive contracts implemented in (military or nonmilitary) practice are of the following (linear) type

$$\pi = \hat{\pi} - s[c - \hat{c}] \quad , \tag{1}$$

where

π is the profit received by the contractor

$\hat{\pi}$ is a target profit

s is a sharing rate $(0 \leq s \leq 1)$

$\hat{c}$ is target cost

c is actual cost.

Formula (1) contains three design parameter $\hat{\pi}, \hat{c}, s$ which are subject to negotiations or competitive bidding. The restriction $0 \leq s \leq 1$ makes sense; values outside $[0,1]$ would stimulate additional strategic considerations with respect to cost reporting. The bigger s the more sensitive is profit as a function of actual cost (compare Fig. 1).

Contracts (1) with $0 < s < 1$ constitute a compromise between two polar cases, namely

- the <u>cost-plus-fixed-fee contract</u> $(s = 0)$ according to which any cost overrun $(c - \hat{c} > 0)$ or cost underrun $(c - \hat{c} < 0)$ is irrelevant for the contractor
- the <u>firm-fixed-price contract</u> $(s = 1)$ according to which the contractor has to bear any cost overrun to the full extent. On the other hand, he also enjoys benefits from any cost underrun to the full extent.

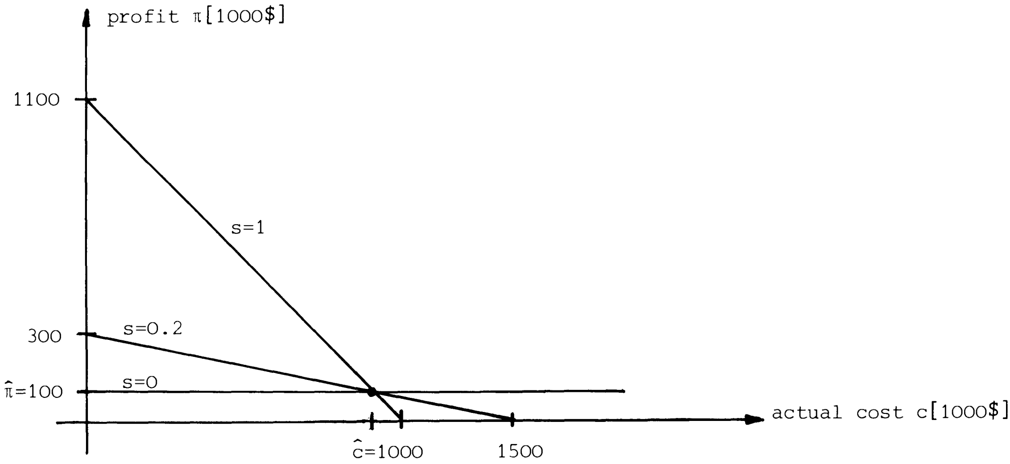


Fig. 1: The effect of different values of the sharing rate s on the profit as function of actual cost (target cost $\hat{c} = 10^6$ $ and target profit $\hat{\pi} = 10^5$ $)

One obstacle to a widespread use of contracts (1) is the difficulty to assess the design parameters $\hat{\pi}, \hat{c}, s$ appropriately. Several approaches have been discussed in the literature: bilateral bargaining, links between design parameters, and competitive bidding with respect to one or several design parameters.

McCall (1970), Baron (1972) and Canes (1975) consider competitive bidding. In the McCall and Baron papers competitive bidding only refers to target cost $\hat{c}$; the sharing rate s is fixed in advance, target profit $\hat{\pi}$ and target cost $\hat{c}$ are linked together (through a given target profit rate). Canes maintains the link between $\hat{\pi}$ and $\hat{c}$ but includes the sharing rate s (in addition to target cost $\hat{c}$) into the competitive bidding process. Links between the design parameters have been considered in nonbidding frameworks too. Scherer (1964a) argues that it is customary for the government to award a higher negotiated target profit $\hat{\pi}$ to contractors who bear a relatively high financial risk - that is, who accept relatively high values of the sharing rate s . Scherer restricts $\hat{\pi}$ to be an increasing quadratic function of the sharing rate s .

Possible follow-on benefits and other dynamic aspects make the assessment problem worse. Follow-on benefits could result from follow-on production contracts or from research and development activities which improve a firm's technical capabilities and enables it to compete for future commercial and military business more sucessfully. Dynamic aspects are treated by Blanning/Kleindorfer/Sankar (1982). They distinguish the development stage from the production stage. During the development stage two or more fixed-price contractors receive funds to design and test a prototype (weapon); at the end of this stage a single contractor is awarded the incentive production contract. Furthermore, the model of Blanning/Kleindorfer/Sankar includes Government's allocation problem between the two stages under consideration.

Contracts (1) with sharing rate s strictly between 0 and 1 are supposed to offer the following advantages:

- The risk sharing provisions makes them acceptable even for contractors or subcontractors who are not able to bear the risk of a firm-fixed-price contract.
- They motivate contractors or subcontractors to control costs more efficiently (compared with cost-plus-fixed-fee contracts).

It should be noted that there is a controversy in the defense related area about Governments's savings due to incentive contracts: Moore (1967) reports that the shift by the Department of Defense from cost-plus-fixed-fee contracts to incentive contracts saved ten cents per dollar expended. According to Fisher (1969), however, there is no empirical evidence for savings. Cost underruns (frequently associated with incentive contracts) are interpreted by him as a consequence of excessive target cost $\hat{c}$. The link between target profit $\hat{\pi}$ and target cost $\hat{c}$ induces a tendency to push up Government's expenditures.

From the theoretical point of view, i.e. judged from normative oriented models, incentive contracts have a clear advantage over cost-plus-fixed-fee contracts. We will reconsider the assessment problem in a typical principal agent setting. Taking into account Fisher's findings we will treat $\hat{\pi},\hat{c}$ and s as independent design parameters. Section 2 describes the model which is highly parsimonious with respect to the number of parameters required to describe cost variations, attitudes towards risk, and the effort function. Since the model includes the remaining economic activities of both the principal and the agent some portfolio aspects are incorporated. In this regard the analysis has some similarities to Berhold (1971). Section 3 focusses on explicit solutions with respect to the design parameters. Typical cases and simplifications are considered in Section 4. Section 5 discusses possible generalizations and modifications.

2. A Parsimoniously Parameterized Modell

If the subcontractor (= agent) correctly meets his contractual obligations, he gets the fee described by (1), i.e.

$$\hat{\pi} - s[c - \hat{c}] \quad . \tag{1}$$

The remaining economic activities of the agent (in terms of profit) will be denoted by the random variable R_a . Actual cost stemming from the correct fulfilment of the contract is also a random variable. This random variable depends on the effort e spent on cost control; we express the dependency and the randomness of c by substituting it through $C(e)$ in (1). By effort e is meant the reduction of the expected cost from μ_c to $\mu_c - e$; the corresponding cost of cost control is $k(e)$. Now we have the random variable relevant for the agent:

$$R_a + \hat{\pi} - s[C(e) - \hat{c}] - k(e) \quad . \tag{2}$$

On the other hand, the contractual payment of the principal (= business firm which offers the contract) is

$$C(e) + \hat{\pi} - s[C(e) - \hat{c}] \quad .$$

The incorporation of the remaining activities (aggregated to the appropriately defined random R_p) results in the random variable

$$R_p - C(e) - \hat{\pi} + s[C(e) - \hat{c}] \tag{3}$$

relevant for the principal.

What is the minimum number of parameters required to analyze such a model? Apart from the three design parameters $\hat{\pi}, \hat{c}, s$ we need at least:

2 parameters to describe the risk attitude of both parties
2 parameters to describe the distribution of actual cost
2 parameters describing the mean values of R_a, R_p
2 parameters describing the variance of R_a, R_p
2 covariances between R_a, R_p and actual cost C
1 parameter to define $k(e)$, the expenditure on cost control measures
1 acceptance level reflecting the agent's market opportunities in case of refusing the contract.

These are 12 additional parameters. In order to keep this minimum number, we will assume all random variables to be normally distributed, all risk aversions to be constant, and the effort function $k(e)$ to be quadratic. More specifically, we will adopt the following notations and assumptions.

α, β	: risk aversions of the agent or principal, respectively
μ_a, μ_p	: expectation of R_a and R_p
σ_a^2, σ_p^2	: variances of R_a and R_p
ρ_a, ρ_p	: correlation coefficient between R_a or R_p with C, resp.
$C(e)$	: distributed according $N(\mu_c - e, \sigma_c)$
$k(e) = \frac{e^2}{2m}$	: effort function.

Note that the assumption of constant risk aversion relieves us from bothering with the agent's and principal's initial wealth. Moreover the combination of constant risk aversion with normally distributed random payments leads to a (μ,σ)-world, and to the well-known and easy-to-handle certainty equivalent (r = degree of risk aversion):

$$\mu - \frac{r}{2}\sigma^2 . \tag{4}$$

This pair of assumptions defines the hybrid approach which turns out to be of utmost convenience in many contexts (compare, for instance, Bamberg (1986), Bamberg/Spremann (1981), Epps (1981), Firchau (1986), Jarrow (1980), Lintner (1970)).

The assumption

$$C(e) \sim N(\mu_c - e, \sigma_c) \tag{5}$$

means that cost controlling efforts have an impact only on expected cost and not on the volatility of cost. Relaxations of this homoscedasticity assumption will be discussed in Section 5.5.

The random variables (2) and (3) may now be substituted by their certainty equivalents. By virtue of (4) and (5) this yields

$$A(\hat{\pi},\hat{c},s,e) = \mu_a + \hat{\pi} - s[\mu_c - e - \hat{c}] - \frac{e^2}{2m} - \frac{\alpha}{2}[\sigma_a^2 - 2s\,\rho_a\sigma_a\sigma_c + s^2\sigma_c^2] \tag{6}$$

with respect to the agent (= subcontractor), and

$$P(\hat{\pi},\hat{c},s,e) = \mu_p - \mu_c + e - \hat{\pi} + s[\mu_c - e - \hat{c}] - \frac{\beta}{2}[\sigma_p^2 - 2(1-s)\rho_p\sigma_p\sigma_c + (1-s)^2\sigma_c^2] \tag{7}$$

with respect to the principal.

3. Explicit Solutions

The principal, concerned with harnessing the cost controlling motive of the agent for his own purposes, has to anticipate the agent's optimal response to a given contract. Therefore, the first step towards the solution of the principal agent problem consists of maximizing the certainty equivalent (6) of the agent.

3.1 Optimal Effort of the Agent

As already notationally indicated, the remaining profits R_a, R_p should be exogenous to the contractual parameters $(\hat{\pi}, \hat{c}, s)$ and to the effort e . Then the agent's certainty equivalent (6) is a quadratic concave function of e . The straightforward maximization yields the optimal effort

$$e_*(\hat{\pi}, \hat{c}, s) = sm \tag{8}$$

with turns out to be independent of the design parameters $\hat{\pi}$ and $\hat{c}$. The optimal effort e_* is an increasing function of the sharing rate s . Formula (3) is in accordance with the intuitive idea that a higher effort results from a higher risk (= greater s) on the part of the agent. The optimal effort is 0 in case of a cost-plus-fixed-fee contract $(s = 0)$, the optimal effort is maximal (= m) in case of a firm-fixed-price-contract $(s = 1)$. Therefore, the effort function parameter m is the maximal reduction of average cost and an indicator for the moral hazard related to the contract. This natural interpretation hints at calibrating m properly. If, for instance, an estimated 100,000 \$ margin exists between the cost levels (corresponding to the least and to the highest effort), m should be fixed at 100,000 \$.

3.2 Pareto Optimal Solutions

Plugging the optimal effort (8) into the agent's and principal's certainty equivalents (6),(7) yields

$$A(\hat{\pi}, \hat{c}, s) = \mu_a + \hat{\pi} - s[\mu_c - sm - \hat{c}] - \frac{s^2 m}{2} - \frac{\alpha}{2}[\sigma_a^2 - 2s\, \rho_a \sigma_a \sigma_c + s^2 \sigma_c^2] = [\hat{\pi} + s\hat{c}] + f_a(s) \tag{9}$$

$$\begin{aligned} P(\hat{\pi}, \hat{c}, s) &= \mu_p - \mu_c + sm - \hat{\pi} + s[\mu_c - sm - \hat{c}] - \frac{\beta}{2}[\sigma_p^2 - 2(1-s)\rho_p \sigma_p \sigma_c + (1-s)^2 \sigma_c^2] \\ &= -[\hat{\pi} + s\hat{c}] + f_p(s) \end{aligned} \tag{10}$$

where

$$\left. \begin{aligned} f_a(s) &= \mu_a - s\mu_c + \frac{s^2 m}{2} - \frac{\alpha}{2}[\sigma_a^2 - 2s\, \rho_a \sigma_a \sigma_c + s^2 \sigma_c^2] \\ f_p(s) &= \mu_p - \mu_c + s\mu_c + sm - s^2 m - \frac{\beta}{2}[\sigma_p^2 - 2(1-s)\rho_p \sigma_p \sigma_c + (1-s)^2 \sigma_c^2] \end{aligned} \right\} \tag{11}$$

Note that the design parameters $\hat{\pi}$ and $\hat{c}$ appear in (9) and (10) only in form of the term $[\hat{\pi} + s\hat{c}]$. From (9) we see that a constant level A of the agent's certainty equivalent can be described by

$$\hat{\pi} + s\hat{c} = A - f_a(s) \ . \tag{12}$$

According to (10), the attainable values of P with respect to the level set (12) are given through

$$P = -A + f_a(s) + f_p(s) \ .$$

This means

$$P + A = f_a(s) + f_p(s) \ . \tag{13}$$

Hence the sum $P + A$ of both certainty equivalents only depends on the sharing parameter s. Clearly, Pareto optimality requires the maximization of

$$f_a(s) + f_p(s) = f(s) \tag{14}$$

subject to $0 \le s \le 1$.

Theorem 1: A contract $(\hat{\pi}, \hat{c}, s)$ is Pareto optimal if and only if s maximizes the function $f(s)$, defined in (11) and (14), subject to the restriction $0 \le s \le 1$.

Fig. 2 illustrates Pareto optimal solutions, i.e. efficient risk sharing arrangements.

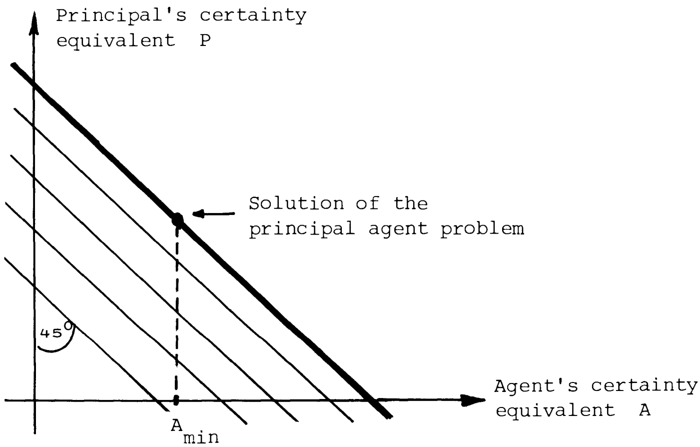


Fig. 2: The different lines correspond to different values of the sharing parameter s ; the bold line corresponds to the optimal value s_* and represents Pareto optimal contracts. The dotted vertical line will be discussed in Section 3.3.

The function $f(s)$, defined in (11) and (14), is

$$f(s) = \mu_a + \mu_p - \mu_c + sm - \frac{s^2 m}{2} - \frac{\alpha}{2}[\sigma_a^2 - 2s\,\rho_a\sigma_a\sigma_c + s^2\sigma_c^2] \tag{15}$$

$$- \frac{\beta}{2}[\sigma_p^2 - 2(1-s)\,\rho_p\sigma_p\sigma_c + (1-s)^2\sigma_c^2] \ .$$

It is concave if both parties are risk averse or risk neutral ($\alpha \geq 0$, $\beta \geq 0$) . As usual, we will rule out risk lovers. Then three cases, illustrated in Fig. 3, must be distinguished.

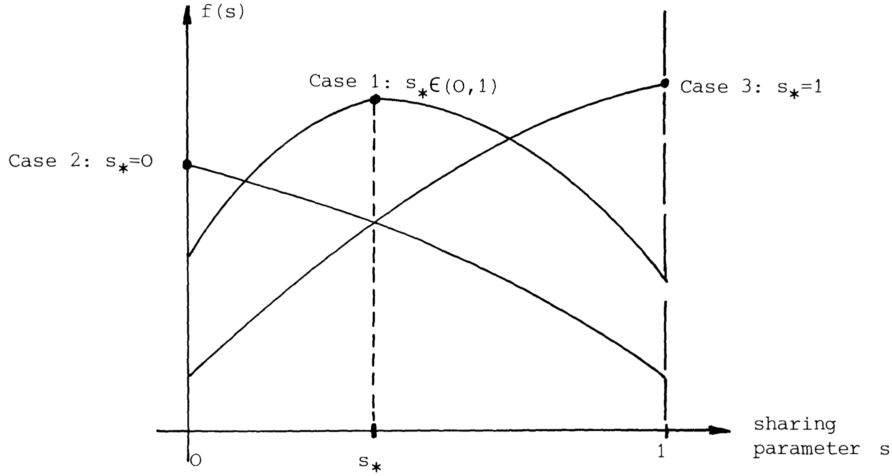


Fig. 3: Shape of the quadratic concave function f(s) defined in (15).

The optimal value s_* , i.e. the value of the sharing parameter s which yields Pareto optimal solutions, is given by

$$s_* = \begin{cases} \dfrac{m + \alpha\, \rho_a \sigma_a \sigma_c - \beta\,[\,\rho_p \sigma_p \sigma_c - \sigma_c^2\,]}{m + (\alpha+\beta)\sigma_c^2} & \text{, if this expression is in } [0,1] \\ 0 & \text{, if the above expression is } < 0 \\ 1 & \text{, if the above expression is } > 1 \end{cases} \qquad (16)$$

3.3 Solution of the Principal Agent Problem

Up to now it was not made use of the agent's acceptance level. If this level will be denoted by A_{min} , we have to solve the problem (compare Fig. 2)

$$\left.\begin{array}{ll} \text{maximize} & P(\hat{\pi},\hat{c},s) \\ \text{subject to} & A(\hat{\pi},\hat{c},s) \geq A_{min} \quad \text{and} \quad 0 \leq s \leq 1 \ . \end{array}\right\} \qquad (17)$$

Obviously, the acceptability restriction may be substituted by the equality

$$A(\hat{\pi},\hat{c},s) = A_{min} ,$$

or - via (12) - by

$$\hat{\pi} + s_*\hat{c} = A_{min} - f_a(s_*) \ . \qquad (18)$$

Equation (18) shows that there is a tradeoff between $\hat{\pi}$ and $\hat{c}$. Theorem 2 summarizes the solution.

Theorem 2: A contract $(\hat{\pi}_*, \hat{c}_*, s_*)$ solves the principal agent problem (17) if and only if
(i) the sharing parameter s_* is given by formula (16) and
(ii) target profit $\hat{\pi}_*$ and target cost $\hat{c}_*$ satisfy equation (18).

4. Discussion of the Optimal Sharing Rate

Formula (16) gives the interior solution and the two corner solutions ($s_* = o$ and $s_* = 1$) with respect to the optimal sharing rate.

In terms of variances and covariances the main case (interior solution) writes as follows

$$s_* = \frac{m + \alpha \ \mathrm{Cov}(R_a, C) - \beta \left[\mathrm{Cov}(R_p, C) - \mathrm{Var}(C) \right]}{m + (\alpha+\beta)\mathrm{Var}(C)} \ . \tag{19}$$

If, for instance, the moral hazard coefficient m becomes bigger and bigger, the optimal contract tends towards a firm-fixed-price contract.

4.1 Effect of Covariances

First of all we will assume that both covariances vanish. In particular this holds if there are no remaining profits or if they can be neglected. Formula (19) then simplifies to

$$s_* = \frac{m + \beta \ \mathrm{Var}(C)}{m + (\alpha+\beta)\mathrm{Var}(C)} \ ,$$

which is always positive (since m , the denominator of the effort function, is positive).

Theorem 3: Let the covariances between the remaining profits (R_a and R_p) and actual cost C be zero. Then the cost-plus-fixed-fee contract cannot be optimal.

Typically, decreasing or even negative correlation offers good diversification opportunities. However, a look on the random variables (2) and (3) shows that the agent faces the risk (measured by variance)

$$\mathrm{Var}\left[R_a - s\,C \right] \tag{20}$$

and that the principal faces the risk

$$\mathrm{Var}\left[R_p - (1-s)\,C \right] \ . \tag{21}$$

The negative sign in (20) and (21) brings it about that increasing covariances (between R_a or R_p and actual cost C) is tantamount to enhanced diversification. Keeping this in mind we should expect an increasing $Cov(R_a,C)$ to correspond to a higher s, since the enhanced diversification enables the agent to bear a higher share of the volatile actual cost. Furthermore, an increasing $Cov(R_p,C)$ should be related to a higher $(1-s)$, since $(1-s)$ is the principal's contractual risky share (compare (21)). But that is exactly what equation (19) tells us:

(i) s_* is an increasing function of $Cov(R_a,C)$
(ii) s_* is a decreasing function of $Cov(R_p,C)$.

4.2 Effect of Risk Attitudes

As we ruled out risk proneness we will focus on the following three cases.

Case A : The agent is risk neutral
Case B : The principal is risk neutral
Case C : Both the principal and the agent are risk neutral.

Case A: *Risk neutral agent* $(\alpha = 0)$ *and risk averse principal* $(\beta > 0)$:

Intuition suggests that the risk neutral agent has to bear the entire contractual risk (i.e. $s_* = 1$) . The simplified equation (19),

$$s_* = \frac{m + \beta \ Var(C) - \beta \ Cov(R_p,C)}{m + \beta \ Var(C)} ,$$

verifies the intuitive result if and only if $Cov(R_p,C) \leq 0$. The already mentioned portfolio argument explains why the agent - despite his risk neutrality - need not bear the entire contractual risk provided that $Cov(R_p,C) > 0$.

Case B: *Risk neutral principal* $(\beta = 0)$ *and risk averse agent* $(\alpha > 0)$:

Again intuition suggests a certain result, namely that the risk neutral principal has to bear the entire contractual risk (i.e. $s_* = 0$) . The simplified equation (19),

$$s_* = \frac{m + \alpha \ Cov(R_a,C)}{m + \alpha \ Var(C)}$$

shows the very limited validity of the intuitive conjecture. The intuitive conjecture $s_* = 0$ is valid if and only if $Cov(R_a,C)$ is negative and sufficiently small (i.e. $Cov(R_a,C) \leq -m/\alpha$) . If $Cov(R_a,C)$ is positive and sufficiently large, even that most counterintuitive result

$s_* = 1$ (the risk averse party bears the entire contractual risk) is possible!

Case C: *Risk neutral principal and risk neutral agent* $(\alpha = \beta = 0)$:

Now equation (19) simplifies to the unambiguous result

$$s_* = 1 \ .$$

Theorem 4: A firm-fixed-price contract is optimal if both the principal and the agent are risk neutral.

5. Some Remarks on the Relaxation of Assumptions

We will touch on some problems stemming from the modification or relaxation of basic assumptions.

5.1 Incorporation of Additional Accounting Costs

The implementation of firm-fixed-price contracts (sharing parameter $s = 1$) is very easy; the principal need not be concerned with actual cost. However, the implementation of contracts with sharing parameter $s < 1$ requires the identification of actual cost c in order to determine the discrepancy between c and target cost $\hat{c}$ and to calculate the fee according to formula (1). The costs arising from the principal's monitoring process and the agent's reporting activities will be termed as monitoring expenditure or as additional accounting costs. If these costs are independent of s (whenever $s < 1$) and are allowed to differ between the two parties, we have:

Additional accounting cost

$$\text{of the agent} = \begin{cases} 0 & , \text{ if } s = 1 \\ \kappa_a & , \text{ if } s < 1 \end{cases} \quad , \quad \text{of the principal} = \begin{cases} 0 & , \text{ if } s = 1 \\ \kappa_p & , \text{ if } s < 1 \end{cases}$$

A check of Section 3 reveals that

- the optimal effort (3) is still valid
- equation (13) must be changed into

$$P + A = \begin{cases} f(s) & , \text{ if } s = 1 \\ f(s) - (\kappa_a + \kappa_p) & , \text{ if } s < 1 \end{cases}$$

- s_* defined in (16) must be compared with $s = 1$ in the following way:

$$\left.\begin{array}{ll} \text{If } f(s_*) - (\kappa_a + \kappa_p) < f(1) & \text{then } s=1 \text{ is optimal (instead of } s_*) \\ \text{If } f(s_*) - (\kappa_a + \kappa_p) \geq f(1) & \text{then } s_* \text{ is still optimal.} \end{array}\right\} \qquad (22)$$

Obviously, sharing values near to one (but < 1) get a comparative disadvantage made precise in (22). The incorporation of additional accounting cost favors firm-fixed-price contracts.

5.2 Arbitrary Effort Functions

We will drop the assumption of a quadratic effort function. Normative reasons suggest $k(e)$ to be monotonously increasing and strictly convex; for the sake of convenience, $k(e)$ is also supposed to be twice differentiable. From (6) we get the implicit representation

$$k'(e_*) = s$$

of the optimal effort e_* . Again, e_* depends on the contractual parameters $(\hat{\pi},\hat{c},s)$ only through s . If we denote the inverse function of k' by h , we get the explicit representation

$$e_* = h(s) \tag{23}$$

of the optimal effort e_* . Since h is monotonously increasing, e_* is an increasing function of the sharing parameter s . Thus the difference $h(1) - h(0)$ between the highest and the lowest effort now plays the role of the moral hazard coefficient m . In general, it seems difficult to derive explicit formulas for the optimal value s_* of the sharing parameter. We will restrict ourselves on the case of risk neutral principal and agent. The maximizing function (15) then boils down to

$$f(s) = \mu_a + \mu_p - \mu_c + h(s) - k[h(s)] \tag{24}$$

which is monotonously increasing since

$$f'(s) = h'(s) - k'[h(s)]h'(s) = h'(s)[1-s] \geq 0 \qquad (0 \leq s \leq 1)$$

(the strict convexity of k entails the positivity of h'). Hence (24) is maximized by the biggest value of the sharing parameter, that is $s_* = 1$.

Theorem 5: Let both the principal and agent be risk neutral and the effort function be an arbitrary (but strictly increasing, strictly convex, twice differentiable) function. Then the firm-fixed-price contract is optimal.

5.3 Nonlinear Contracts

Nonlinearities arise, for instance, from profit ceilings or asymmetries between cost overruns and cost underruns. Two types of nonlinear contracts, the fixed-price-incentive-fee contract and the cost-plus-incentive-fee contract has been mentioned in the defense related literature and will be explained below. The nonlinearities make it extremly difficult to obtain explicit solutions. The fact that nonlinear transformations of normally distributed random variables are no longer normally distributed leads to the abandonment of the convenient hybrid model.

Fixed-price-incentive-fee contracts are characterized by the following provisions:

- the principal shares all cost underruns with the agent according to formula (1).
- the agent has to absorb all cost overruns.

More generally, the provisions can specify a cost limit $\hat{\hat{c}}$ ($\geq \hat{c}$) up to which the agent's profit satisfies (1) and beyond which the agent has to absorb all additional costs:

$$\pi = \begin{cases} \hat{\pi} - s(c - \hat{c}) & , \text{ if } c \leq \hat{\hat{c}} \\ \hat{\pi} - (c - \hat{c}) & , \text{ if } c > \hat{\hat{c}} \end{cases} \tag{25}$$

The term fixed-price-incentive-fee contract stems from the property of contract (25) to behave like a firm-fixed-price contract if actual cost exceeds the limit $\hat{\hat{c}}$. Fig. 4 illustrates formula (25). Fixed-price-incentive-fee contracts have been discussed in some detail by Cross (1968).

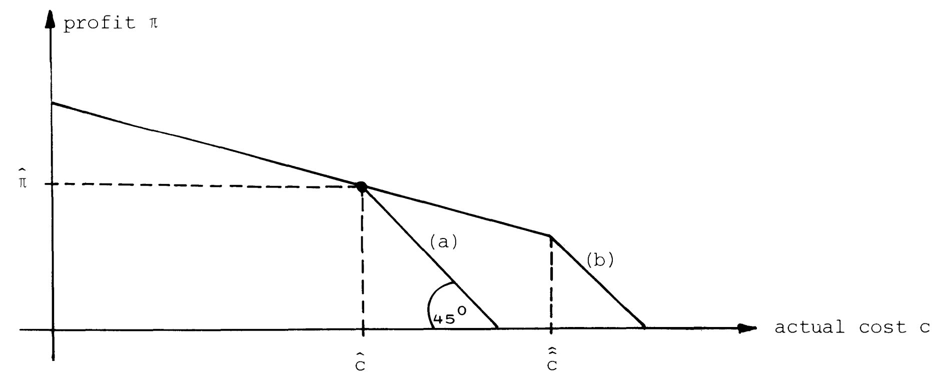


Fig. 4: Profit according to a fixed-price-incentive-fee contract under wich the agent must absorb all cost overruns (case (a)) or cost overruns beyond the prespecified level $\hat{\hat{c}}$ (case b)).

Cost-plus-incentive-fee contracts provide lower and upper bounds with respect to profit. Within these limits profit is given by formula (1):

$$\pi = \begin{cases} \hat{\pi} - s(c-\hat{c}) & \text{, if this expression is in } [\pi_{min}, \pi_{max}] \\ \pi_{min} & \text{, if " " is } < \pi_{min} \\ \pi_{max} & \text{, if " " is } > \pi_{max} \end{cases}$$

Fig. 5 gives the shape of the corresponding cost-profit relationship. Obviously, the term cost-plus-incentive-fee contract stems from the behavior like a cost-plus-fixed-fee contract outside the range $[\pi_{min}, \pi_{max}]$.

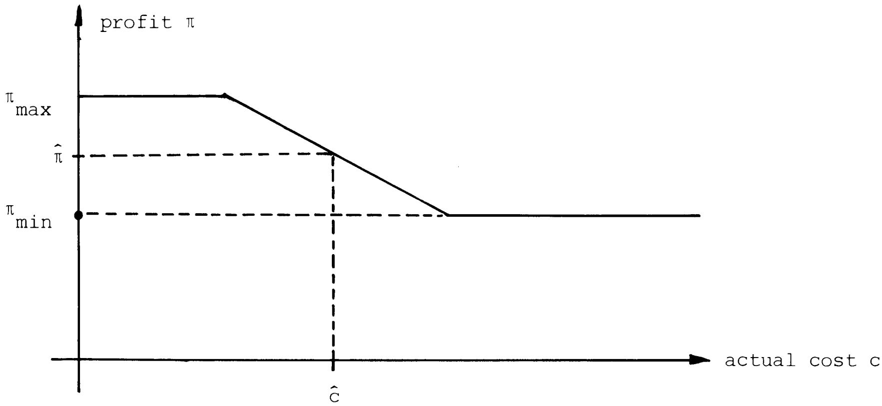


Fig. 5: Cost-profit relationship induced by a cost-plus-incentive-fee contract.

Sometimes cost-plus-incentive-fee contracts are furnished with additional incentive provisions, namely bonuses or penalties related to product quality characteristics or time of completion.

5.4 Heterogeneous Expectations

Up to now we assumed that both the principal and the agent base their decisions on the same (joint) probability distribution of $C(e), R_a, R_p$. Following portfolio terminology, by heterogeneous expectations is meant a discrepancy between the probability distributions assessed by principal and agent. One could even think of a third distribution (the true distribution). Furthermore, the assessment of the utility functions and the effort function could be subject to heterogeneity. The hybrid model is sufficiently easy to handle heterogeneous expectations (with respect to the probability distributions). Nevertheless we leave it out of account because numerous subcases are to distinguish.

In passing, consider for instance the agent's level set defined in (12). Heterogeneous expectations entail the distinction between the level set calculated by means of the agent's distribution and the level set perceived by the principal (i.e. calculated by means of the principal's distribution). The concept of Pareto optimality looses its unambiguity. One has to specify the distributions from which the different (A,P)-pairs have been calculated etc.

5.5 Effort Dependent Cost Variance

The impacts of different effort levels e on the distributional parameters of actual cost have been assumed as follows (compare (5)):

$$\mu_c(e) = \mu_c - e \quad ; \quad \sigma_c(e) = \sigma_c \; . \tag{26}$$

We will briefly discuss three scenarios different from (26); all other basic assumptions will be maintained.

Scenario 1: $\mu_c(e) = \mu_c - e \; ; \; \sigma_c(e) = \sigma_c - de \quad (d \geq 0)$

According to scenario 1, cost controlling efforts reduce both expected value and standard deviation of actual cost; the previous case (26) corresponds to $d = 0$.

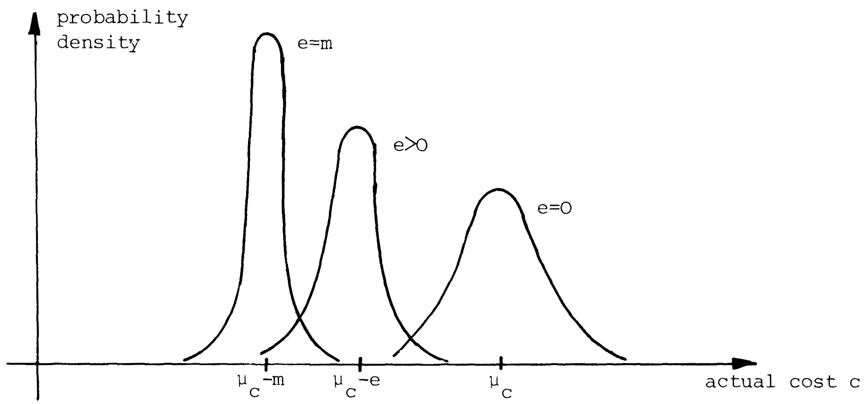


Fig. 6: Probability density of actual cost resulting from three different levels (m,e, and 0) of effort under scenario 1; the higher the effort the more concentrated is the probability distribution.

The optimal effort now turns out to be

$$e_*(s) = sm \, \frac{1 - \alpha \, \rho_a \sigma_a \, d + \alpha \, \sigma_c \, d \, s}{1 + m \, \alpha \, d^2 \, s^2} \; . \tag{27}$$

It is not necessarily increasing in the sharing parameter s , the intuitive argument "higher contractual risk entails higher effort" is not always true. However, the case of risk neutral agents $(\alpha = 0)$

leads us back to the increasing function (8):

$$e_*(s) = sm \ .$$

If in addition, the principal displays risk neutrality (= case C of Section 4.2) the optimal contract is a firm-fixed-price contract.

Scenario 2: $\mu_c(e) = \mu_c$; $\sigma_c(e) = \sigma_c - de$ $(d \geq 0)$.

According to scenario 2, cost controlling efforts have an impact only on cost variance. The optimal effort turns out to be

$$e_*(s) = sm\,\alpha d \frac{s\sigma_c - \rho_a \sigma_a}{1 + m\,\alpha\, d^2 s^2}$$

Let the agent be risk neutral $(\alpha = 0)$. In this instance scenario 2 is useless since it is impossible to motivate the agent for a higher effort than $e = 0$.

Scenario 3: $\mu_c(e) = \mu_c + e$; $\sigma_c(e) = \sigma_c - de$ $(d > 0)$

Scenario 3 is inspired by the fact that most variance reducing activities are cost-intensive: hedging of various kinds, stockpiling of raw material or trained staff, using futures markets instead of relying on spot markets etc. Therefore, it is not unreasonable to assume that one has to pay for reduced cost uncertainty by increasing expected cost. Unfortunately, scenario 3 shares with scenario 2 the property to break down when risk neutral agents are involved. The optimal effort of such an agent is

$$e_*(s) = -sm \ ;$$

the higher the contractual risk, the lower the effort! This seemingly paradox result can easily be explained: The risk neutral agent does not bother about variance reductions, he is only interested in the reduction of expected actual cost. Therefore, he tends to negative e-values. The example also casts some doubt on the term "effort" to mark the parameter e in general scenarios.

References

Bamberg, G. (1986): The Hybrid Model and Related Approaches to Capital Market Equilibria, in Bamberg, G.; Spremann, K. (eds.): Capital Market Equilibria, Springer Publ. Comp., Berlin-Heidelberg-New York-Tokyo, 7-54

Bamberg, G.; Spremann, K. (1981): Implications of Constant Risk Aversion, Zeitschrift für Operations Research 25, 205-224

Baron, D.P. (1972): Incentive Contracts and Competitive Bidding, American Economic Review 62, 384-394

Berhold, M. (1971): A Theory of Linear Profit-Sharing Incentives, Quarterly Journal of Economics 85, 460-482

Blanning, R.W.; Kleindorfer, P.R.; Sankar, C.S. (1982): A Theory of Multistage Contractual Incentives with Application to Design-to-Cost, Naval Research Logistics Quarterly 29, 1-18

Canes , M.E. (1975): The Simple Economics of Incentive Contracting: Note, American Economic Review 65, 478-483

Cross, J.G. (1968): A Reappraisal of Cost Incentives in Defense Contracts, Western Economic Journal 6, 205-225

Cummins, M.T. (1977): Incentive Contracting for National Defense: A Problem of Optimal Risk Sharing, The Bell Journal of Economics 8, 168-185

Epps, T.W. (1981): Necessary and Sufficient Conditions for the Mean-Variance Portfolio Model with Constant Risk Aversion, Journal of Financial and Quantitative Analysis 16, 169-176

Firchau, V. (1986): Portfolio Decisions and Capital Market Equilibria Under Incomplete Information, in Bamberg, G.; Spremann, K. (eds.): Capital Market Equilibria, Springer Publ. Comp., Berlin-Heidelberg-New York-Tokyo, 55-78

Fisher, I.N. (1969): An Evaluation of Incentive Contracting Experience, Naval Research Logistics Quarterly 16, 63-83

Fox, J.R. (1974): Arming America: How the U.S. Buys Weapons, Harvard Univ. Press, Cambridge (Mass.)

Jarrow, R. (1980): Heterogeneous Expectations, Restrictions on Short Sales, and Equilibrium Asset Prices, Journal of Finance 35, 1105-1113

Kawasaki, S.; McMillan, J. (1985): The Design of Contracts: Evidence from Japanese Subcontracting, Arbeitspapiere zur Strukturanalyse Nr. 31, Univ. Augsburg

Kleindorfer, P.R.; Sertel, M.R. (1979): Profit-Maximizing Design of Enterprises Through Incentives, Journal of Economic Theory 20, 318-339

Krahnen, J.-P.; Schmidt, R.H.; Terberger, E. (1986): Der ökonomische Wert von Flexibilität und Bindung, 253-285 in Ballwieser, W.; Berger, K.-H. (Hrsg.): Information und Wirtschaftlichkeit, Wiesbaden

Lintner, J. (1970): The Market Price of Risk, Size of Market, and Investor's Risk Aversion, Review of Economic Studies 52, 87-99

Macdonald, G.M. (1984): New Directions in the Economic Theory of Agency, Canadian Economics Association 17, 415-440

McCall, J.J. (1970): The Simple Economics of Incentive Contracting, American Economic Review 60, 837-846

Moore, F.T. (1967): Incentive Contracts, in Enke, S. (ed.): Defense Economics, Prentice/Hall, Englewood Cliffs, Ch. 12

Peck, M.J.; Scherer, F.M. (1962): The Weapons Acquisition Process: An Economic Analysis, Harvard Univ. Press, Cambridge (Mass.)

Scherer, F.M. (1964a): The Theory of Contractual Incentives for Cost Reduction, Quarterly Journal of Economics 78, 257-280

Scherer, F.M. (1964b): The Weapons Acquisition Process: Economic Incentives, Harvard Univ. Press, Cambridge (Mass.)

Shavell, S. (1979): Risk Sharing and Incentives in the Principal and Agent Relationship, Bell Journal of Economics 10, 55-73

Tashjian, M.T. (1974): Implementation of the Design-to Cost Concept from the Contractual Point of View, Defense Management Journal 10, 8-17

Williamson, O.E. (1967): The Economics of Defense Contracting: Incentives and Performance, in McKean, R.N. (ed.): Issues in Defense Economics, Columbia Univ. Press, New York, 217-256.

Section 2 Information and Incentives

Information Systems for Principal-Agent Problems

Volker Firchau

Summary: For the basic model of decision theory it is well-known how to use and evaluate additional information about the unknown parameter which determines the distribution of the outcome. Principal-agent problems are more complicated. Two decision makers and two types of unknown quantities have to be distinguished. This leads to several variants when analyzing the influence of additional information for principal-agent problems. The paper summarizes some known results, presents a few new ones and gives an outlook to further research.

Table of contents

1. Introduction

Agency theory, although not quite new, is one part of decision theory which is characterized by a dynamic development at the present time. The well-known basic model of decision theory is related to a situation in which one person - the so called decision maker - has to make a decision as well as to bear the consequences. This assumption is often unrealistic. Therefore, in the agency theory the identity between the decision maker (the agent) and the usufructuary of the decision (the principal) is abolished. Typical examples are given in Fig. 1.

If the principal cannot influence the decision of the agent, a usual decision problem results for the agent who tries to minimize his/her effort ignoring the actual consequences. In the principal-agent model, on the other hand, a compensation is admitted paied by the principal to the agent. This payment may depend on the result of the decision for the principal and/or on other information and is, therefore, a function. The compensation function shall induce the agent to act according

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

PRINCIPAL	AGENT	DECISION
1. society	firm	provisions against enviromental damages
2. owner of a firm	manager	management decisions strategic decisions
3. manager of a firm	subordinate	concrete decisions, jobs
4. patient	doctor	medical treatment
5. client	lawyer	consultation
6. insurer	insurant	insured activity

Fig. 1 : Examples for principal-agent problems

to the principal's interests. The determination of an optimal compensation function is the main problem of the principal-agent theory. Beside the mentioned incentive effect the compensation function shall and can induce a satisfactory risk sharing between the principal and the agent, especially for the case of different attitudes towards risk. For the examples given in Fig. 1 compensations may be:

- fiscal incentives, subsidies or fines (example 1)
- more or less performance-based fees (example 2 - 5)
- boni, premium reduction in the case of experience rating (example 6).

Therefore, each of the two involved persons has to make a decision: the principal about the compensation and the agent about the real decision. Corresponding to the most realistic situations, it is assumed that the agent can accept an offered compensation contract or can reject it. A principal-agent problem turns out to be a special case of a dynamic two-person game. First, the two players cooperatively determine the outcome functions and then play the game in a noncooperative way. This is a complete symmetric situation and, indeed, there are examples in which it is not obvious who is the principal and who the agent.

One foundation of the principal-agent theory certainly is Herbert A. Simon's paper of 1951 which contains especially many ideas corresponding to information analysis concepts (see Mattessich (1984)). The topic of the present paper is to give a survey about the possibilities to introduce information concepts in the principal-agent theory, to summarize some important known results, to present a few new ones and to give an outlook to further research.

In the 'classical' decision theory, the influence of additional information about the unknown parameter on the decision-making is analyzed. Information-dependent decision, i.e. decision functions, are considered. How can this extension be transferred to the principal-agent model? Here, the situation is more complicated. Not only two decisions are to be made but also two unknown quantities are to be distinguished: firstly, the decision parameter unknown to the principal and the agent and, secondly, the decision of the agent which is unknown to the principal. This leads to several variants of the model some of which are discussed in the following. For simplicity, some technical assumptions although necessary are omitted.

2. Decision Problems with Additional Information

Before the principal-agent model is analyzed it is appropriate to recapitulate some results for one-person decision problems. The following simple model is considered:

A decision maker has to choose a decision d from a set D of possible decisions. The result x depends on d and an unknown parameter $\vartheta : x = x(\vartheta, d)$. The parameter ϑ is an element of a finite set Θ . If the decision maker has a prior distribution p for ϑ and a utility function u he/she has to solve the following optimization problem:

$$\max_{d \in D} \sum_{\vartheta \in \Theta} u[x(\vartheta,d)]p(\vartheta) =: c \tag{1}$$

The temporal succession of the process is illustrated by the following diagram:

$$d \in D \rightarrow \vartheta \in \Theta \rightarrow x(\vartheta,d)$$

As ϑ is unknown, additional information about ϑ are desirable from a pragmatic point of view. If the realization y of a random variable Y whose distribution depends on ϑ can be observed, the following optimization problem results:

$$\max_{\delta \in \Delta_Y} \sum_{y,\vartheta} u[x(\vartheta,\delta(y))]p_\vartheta(y)p(\vartheta) =: c(Y) \tag{2}$$

Δ_Y is a set of admitted decision functions which are mappings on the finite set $\{y\}$ of possible outcomes of Y to D :

$$\Delta_Y \subset \{\delta : \{y\} \rightarrow D\}$$

Again, it is instructive to observe the process in time:

$$Y \in \mathcal{Y} \rightarrow \delta \in \Delta_Y \rightarrow \vartheta \in \Theta \rightarrow Y = y \rightarrow x(\vartheta,\delta(y))$$

At first, one information system Y (perhaps 'no information') has to be chosen from a set $\mathcal{Y}$ of possible information systems.

Three basic questions arise:

- What is the optimal δ ?
- What is the best (costless) information system?
- Which information costs are tolerable?

The first point is the topic of the theory of Bayes decision functions. The second question is partly answered by the two statements:

- $c(Y) \geq c$ if $D \subseteq \Delta_Y$, i.e. if the constant decision functions are elements of Δ_Y .
- Blackwell's theorem: Let $Y,Z \in \mathcal{Y}$, then $C(Y) \geq C(Z)$ for all attachable decision problems if and only if there exists a Markov kernel $p(z|y)$ with

$$p_\vartheta(z) = \sum_y p(z|y)p_\vartheta(y) \tag{3}$$

(see Blackwell, Girshick (1966); here, a technical assumption is necessary to make sure that Δ_Y is not too 'small').

Roughly spoken, the results are: Information is always better than no one and randomization never pays.

The maximum tolerable information costs are named expected value of information (EVI) which can be explicitly determined only under additional assumptions. For an exponential utility function:

$$u(x) = -e^{-\alpha x} \quad (\alpha > 0) \tag{4}$$

LaValle (1963) showed

$$EVI = u^{-1}(c(Y)) - u^{-1}(c)$$

which is not negative if $D \subseteq \Delta_Y$.

3. Principal-Agent Problems

The principal-agent problem corresponding to (1) is

$$\max_{s \in S_X} \sum_{\vartheta \in \Theta} u_P[x(\vartheta,d) - s(x(\vartheta,d))]p(\vartheta) =: c_P \tag{5}$$

subject to

$$\sum_{\vartheta \in \Theta} u_A[s(x(\vartheta,d))]p(\vartheta) - v(d)$$

$$= \max_{d' \in D} \{ \sum_{\vartheta \in \Theta} u_A[s(x(\vartheta,d'))] - v(d')\} \geq \bar{u} \tag{6}$$

where

u_P, u_A are the utility functions of the principal and the agent, respectively

S_X is the set of considered compensation functions defined on the set of possible outcomes $\{x\}$:
$S_X \subset \{s : \{x\} \to \mathbb{R}\}$

v is the disutility function of the agent

$\bar{u}$ is the reservation level of the agent.

If the principal offers s the agent will choose d according to (6) if the reservation level $\bar{u}$ is reached. Otherwise, he/she will reject the contract. For the optimal compensation function often the equality holds in (6). The model is only applicable if the principal knows all data while the agent only has to know his/her decision problem (6). It should be mentioned that (5),(6) is a very simple version of the principal-agent problem. Possible extensions are obvious. The process is put through in the following way:

$$s \in S_X \to d \in D \to \vartheta \in \Theta \to x(\vartheta,d) \to s(x(\vartheta,d))$$

Two types of information systems can be considered: information about ϑ and information about d . The first kind called here parameter information corresponds to the information systems considered in sec. 2. Again, an information system is a random variable Y whose distribution depends on ϑ . Three subcases can be distinguished ($S_{X,Y} \subset \{s : \{x\} \times \{y\} \to \mathbb{R}\}$)

- parameter information for the principal:

$$Y \in \mathcal{Y} \to s \in S_{X,Y} \to d \in D \to \vartheta \in \Theta \to Y = y \to x(\vartheta,d) \to s(x(\vartheta,d),y)$$

- parameter information for the agent:

$$Y \in \mathcal{Y} \to s \in S_X \to \delta \in \Delta_Y \to \vartheta \in \Theta \to Y = y \to x(\vartheta,\delta(y)) \to s(x(\vartheta,\delta(y)))$$

- parameter information for the principal and the agent:

$$Y \in \mathcal{Y} \to s \in S_{X,Y} \to \delta \in \Delta_Y \to \vartheta \in \Theta \to Y = y \to x(\vartheta,\delta(y)) \to s(x(\vartheta,\delta(y)),y)$$

It is assumed that each involved person always knows if the other will be informed and that both have homogeneous expectations. For example, in the first case the agent knows when making his/her decision that the principal will observe Y . This is obvious because the agent knows s which depends on the realization of Y .

The second type of information is called agency-information (see Gjesdal (1980, 1982), Mattessich (1984)). The distribution of Y now depends on ϑ and d . Obviously, such information can only be observed after the agent made his/her decision. Agency-information, therefore, is only relevant for the principal. So, agency-information turns out to be a formal generalization of parameter information for the principal with identical temporal sucession.

In the following, the different types of information are analyzed in more detail.

4. Parameter Information

Firstly parameter information for the principal is considered. The model corresponding to (2) is

$$\max_{s \in S_{X,Y}} \sum_{y,\vartheta} u_P[x(\vartheta,d) - s(x(\vartheta,d),y)]p_\vartheta(y)p(\vartheta) =: c_P^P(Y) \tag{7}$$

subject to

$$\sum_{y,\vartheta} u_A[s(x(\vartheta,d),y)]p_\vartheta'(y)p(\vartheta) - v(d) = \max_{d' \in D} \{ \sum_{y,\vartheta} u_A[s(x(\vartheta,d'),y)]p_\vartheta(y)p(\vartheta) - v(d')\} \geq \geq \bar{u} \tag{8}$$

Examples for such an information system are (compare Fig. 1):

- measurements of emissions in the neighbourhood of the firm (example 1)
- reports of independent experts about the 'situation' (examples 2 - 6).

The agent's level of information is unchanged. Therefore, all compensation functions accepted without information will be accepted further on and lead to the same decision:

$$c_P^P(Y) \geq c_P \quad \text{if} \quad S_X \subsetneq S_{X,Y}$$

If u_P is exponential (see(4)) it holds again

$$EVI = u_P^{-1}(c_P^P(Y)) - u_P^{-1}(c_P)$$

which is not negative if $S_X \subsetneq S_{X,Y}$.

Whereas the principal normally is better off, in (8) as in (6) often the equality sign holds, i.e. the situation has not changed for the agent.

The problem is not analyzed in more detail because it is only a special case of agency information. But it should be mentioned that Holmström (1979) proved a statement corresponding to Blackwell's theorem under some additional assumptions and for the case that Z is the 'no information' system.

The next point is parameter information for the agent, perhaps the most interesting model. Such information systems, also named 'differential information' are briefly considered by Harris, Raviv (1972), Holmström (1979) and Gjesdal (1980). A very important paper to the subject is due to Christensen (1981), who showed by an example that such information can have a negative value for the principal. A modification of his example is given below. The formal description of the model is:

$$\max_{s\in S_X} \sum_{y,\vartheta} u_P[x(\vartheta,\delta(y)) - s(x(\vartheta,\delta(y)))]p_\vartheta(y)p(\vartheta) =: c_P^A(Y) \tag{9}$$

subject to

$$\sum_{y,\vartheta} [u_A[s(x(\vartheta,\delta(y)))] - v(\delta(y))]p_\vartheta(y)p(\vartheta)$$

$$= \max_{\delta'\in\Delta_Y} \{ \sum_{y,\vartheta} [u_A[s(x(\vartheta,\delta'(y)))] - v(\delta'(y))]p_\vartheta(y)p(\vartheta)\} \geq \bar{u} \tag{10}$$

Exemples for such information systems are internal training, visits of courses, seminars and conferences. It is to be hoped that such activities lead to a higher productivity.

Who is better off in this new situation? One might expect that the position of the agent has been improved if $D \subsetneq \Delta_Y$ as a result of higher decision flexibility. But there are also advantages for the principal. The agent may accept compensation functions which are inadmissible without additional information. On the other hand, an informed agent can better evaluate the result x of a decision. This could lead to a lower effort of the agent to the principal's disadvantage, particularly, if the parameter ϑ has a favourable value. The following examples show this contrary effects for the case of a perfect information system Y that reports the true parameter ϑ.

<u>Examples:</u>

a) Let u_A, u_P be strictly concave and monotone and the disutility function be constant: $v(d) = v$. For an estimation problem $(\theta = D)$ with result function:

$$x(\vartheta,d') = \begin{cases} 1 & \text{if } \vartheta = d' \\ 0 & \text{if } \vartheta \neq d' \end{cases}$$

obviously, the following compensation function

$$s(x) = \begin{cases} u_A^{-1}(\bar{u} + v) & \text{if } x = 1 \\ u_A^{-1}(\bar{u} + v) - 1 & \text{if } x = 0 \end{cases}$$

has to be accepted by the agent if he/she is perfectly informed about ϑ . The corresponding optimal decision function is $\delta^*(y) = \vartheta$ which leads to the so-called 'first-best' solution: $c_A^P(y) \geq c_P$. The equality sign only holds in trivial cases as $|\Theta| = 1$.

b) (Compare Christensen (1981))
Let $D = \mathbb{R}$, $\Theta = \{0,1\}$, $p \in (0,1)$, $\vartheta \sim B(1,p)$, $\bar{u} > 0$, $v(d') = \max\{d'^2, 0\}$, $x(\vartheta, d') = \vartheta + d'$, $u_p(x) = x$, $u_A(s) = 2\sqrt{s}$ (S_X may contain only non-negative functions) and $d^* \in (0,1)$ be the unique solution of $d^*(\bar{u} + d^{*2}) = 1$. It is easy to show that the compensation function

$$s(x) := \begin{cases} \frac{1}{4d^{*2}} & \text{if} \quad x \geq d^* \\ 0 & \text{if} \quad x < d^* \end{cases}$$

induces the decision d^* by the agent and together leads to the 'first-best' solution, if $d^{*3} < 1-p$. This is for example fulfilled if $p = 7/8$ and $\bar{u} = 63/16$, which result in $d^* = 1/4$.

Now the agent may observe perfect information, i.e. the true parameter ϑ . If Δ_Y is not too small ($\Delta_Y = \{\delta' : \{y\} \to D\} \mathrel{\hat{=}} \mathbb{R}^2$ for example) than any compensation function leads to a not constant decision function for the agent and to a situation which is worse than the first-best solution for the principal. This is true for each compensation function and, therefore, also for the best one: $c_P^A(Y) < c_P$.

The last example shows that the Blackwell condition (3) is not sufficient for an information system to be preferred against another one. It seems that for the effect of information it is decisive whether the principal-agent relation is more cooperative or not. A typical cooperative situation is described in the following theorem.

<u>Theorem 1:</u> Let $v(d) = v$, Y be perfect, $\Delta_Y = \{\delta' : \{y\} \to D\}$, and the functions $u_A(x), u_P(x), s^*(x)$ and $x - s^*(x)$ be strictly monoton, where s^* is the optimal compensation function without information, then $c_P^A(Y) \geq c_P$.

The proof is obvious as already s^* does not lead to a deterioration for the principal. This is an interesting robustness property of s^* . If the principal chooses s^* then he/she does not have to be afraid of additional perfect information for the agent.

If u_P is exponential (see (4)) then again

$$EVI = u_P^{-1}(c_P^A(Y)) - u_P^{-1}(c_P) \tag{11}$$

results.

If the EVI is positive the principal is interested in an informed investor. On the other hand, the agent often will not be in a better position with information especially if the equality sign holds in (10). Among others, the following two incentives seem reasonable to induce the agent to gather information.

- The principal does not change the compensation function. This make sense in situations as described in theorem 1 when both will be better off.
- The agent can enforce a higher reservation level as a consequence of his/her improved qualification.

The second possibility shall be discussed in some detail. The increase of the reservation level deteriorates the result for the principal. What is the maximum tolerable reservation level? The following theroem gives the answer for a special case.

Theorem 2: Let u_A, u_P be exponential, S_X be closed under addition of constants:

$$s \in S_X, \ a \in \mathbb{R} \Rightarrow s + a \in S_X$$

and $\bar{u}$ be so that for the optimal compensation function s^* the equality sign holds in (10), then an increase Δu of the reservation level of the informed agent is tolerable by the principal (i.e. $c_P \leq c_P^A(Y)$ for the new reservation level) as long as

$$\Delta u \leq (e^{-\alpha(EVI-K(Y))} - 1)(Ev(\delta^*(Y)) + \bar{u}) \qquad (12)$$

where

- δ^* is the optimal decision function of the agent (for the optimal compensation function, see (10)), which is the same for each reservation level for which a solution of (9),(10) exists
- α is the risk aversion of the agent (see (4)):

  $$u_A(s) = -e^{-\alpha s}$$

- EVI is the expected value of information if $\bar{u}$ is the reservation level (see (11))
- $K(Y)$ are the information costs.

Proof: It is obvious that

$$s^{**} = s^* + u_A^{-1}(Ev(\delta^*) + \bar{u} + \Delta u) - u_A^{-1}(Ev(\delta^*) + \bar{u})$$

is the optimal compensation function for the reservation level $\bar{u} + \Delta u$ if $Ev(\delta^*) + \bar{u} + \Delta u < 0$. This is fulfilled if (12) holds as

$E(v(\delta^*)) + \bar{u} < 0$ (see (10)). The optimal corresponding decision function is again δ^*, and the assertion follows immediately from (9).

Finally the same parameter information for the principal and the agent is considered:

$$\max_{s \in S_{X,Y}} \sum_{y,\vartheta} u_P[x(\vartheta,\delta(y)) - s(x(\vartheta,\delta(y)),y)]p_\vartheta(y)p(\vartheta) =: c_P^{P,A}(Y) \tag{13}$$

subject to

$$\sum_{y,\vartheta} [u_A[s(x(\vartheta,\delta(y)),y)] - v(\delta(y))]p_\vartheta(y)p(\vartheta)$$
$$= \max_{\delta' \in \Delta_Y} \left\{ \sum_{y,\vartheta} [u_A(s(x(\vartheta,\delta'(y)),y)] - v(\delta'(y))]p_\vartheta(y)p(\vartheta) \right\} \geq \bar{u} \tag{14}$$

Again the principal is not forced to use the information:

$$c_P^{P,A}(Y) \geq c_P^A(Y) \quad \text{if} \quad S_X \subsetneq S_{X,Y}$$

If the agent gets information it is no disadvantage if the principal is informed too. If $c_P^A(Y) \geq c_P$, i.e. in situations as described in theorem 1, such information is favourable at all. For Christensen's example, it es easy to show that $c_P^{P,A}(Y) = c_P$, i.e. the negative effect is compensated. In reality, often one of the involved persons is informed first and will then inform the other one more or less correctly. Such communication models are considered by Christensen (1981) and Penno (1985).

5. Agency Information

In contrast to the cases treated sec.4, there are several publications related to agency information, for example: Christensen (1981), Gjesdal (1980,1982), Harris/Raviv (1972), Holmström (1979), Kanodia (1985), Mattessich (1984), Penno (1985), Ramakrishnan/Thakor (1982), Shavell (1979), and Singh (1985). Most of the papers analyze special types of agency information as a (disturbed) observation of the decision made by the agent. A formal description of the model is

$$\max_{s \in S_{X,Y}} \sum_{y,\vartheta} u_P[x(\vartheta,d) - s(x(\vartheta,d),y)]p_{\vartheta,d}(y)p(\vartheta) =: c_P^{AI}(Y) \tag{15}$$

subject to

$$\sum_{y,\vartheta} u_A[s(x(\vartheta,d),y)]p_{\vartheta,d}(y)p(\vartheta) - v(d)$$
$$= \max_{d' \in D} \left\{ \sum_{y,\vartheta} u_A[s(x(\vartheta,d'),y)]p_{\vartheta,d'}(y)p(\vartheta) - v(d') \right\} \geq \bar{u} \tag{16}$$

A comparison of (7),(8) with (15),(16) shows, that agency information is, as mentioned, a generalization of parameter information for the principal. The interest for agency information can perhaps be explained by the fact that there is no analogon for agency information in the basic model of decision theory (see sec. 2) where the decision maker of course is perfectly informed about his/her decision. Again, it follows immediately:

$$c_P^{AI}(Y) \geq c_P \quad \text{if} \quad \mathcal{S}_X \subsetneq \mathcal{S}_{X,Y}$$

The following theorems show that Blackwell's theorem is only valid in one direction. Some technical assumptions about the sets of possible compensation functions are omitted. They do not have to be too 'small'.

Theorem 3: (Harris, Raviv 1972)
Let Y be perfect for ϑ amd Z be perfect for ϑ and d, then:

$$c_P^{AI}(Y) = c_P^{AI}(Z) \quad \text{if} \quad \mathcal{S}_{X,Y} \subsetneq \mathcal{S}_{X,Z}$$

If, in addition, the agent is risk-neutral and the principal risk-averse, then:

$$c_P^{AI}(Z) = c_P \qquad \text{if} \quad \mathcal{S}_X \subsetneq \mathcal{S}_{X,Z}$$

Theorem 4: (Gjesdal 1982)
Let u_A, u_P to concave and u_A strictly monotone, then $c_P^{AI}(Y) \geq c_P^{AI}(Z)$ (for each attachable principal-agent problem) if there is a Markov kernel $p(z|y)$ with:

$$p_{\vartheta,d}(z) = \sum_y p(z|y) p_{\vartheta,d}(y)$$

Theorem 4 correspond to one direction of Blackwell's theorem. The first part of theorem 3 is a counter example for the other direction: there is no Markov kernel $p(z|y)$. The second part of theorem 3 shows that there is no moral hazard problem if the agent is risk-neutral.

6. Conclusion

Instead of a summary some interesting open problems are mentioned:

- What are conditions which are sufficient for

  $c_P^{A}(Y) \geq c_P$, $c_P^{P,A}(Y) \geq c_P^{P}(Y)$, $c_P^{P,A}(Y) \geq c_P$?

- Are there interesting problems for which qualitative results can bei derived?

How can the results be used for an integrated planning of hierarchy, incentive system and information flow within an organization?

The papers of Mirrless (1976) and Singh (1985) show some promise with respect to the last question.

References

Blackwell, D.; Girshick, M.A. (1966) Theory of Games and Statistical Decisions, 5. ed., John Wiley & Sons, Inc., New York

Christensen, J. (1981): Communication in Agencies, The Bell Journal of Economics, 12, 661-674

Gjesdal, F. (1980): Information and Incentives: The Agency Information Problem, Discussion Paper 12/80, Norwegian School of Economics and Business Administration

Gjesdal, F. (1982): Information and Incentives: The Agency Information Problem, Review of Economic Studies XLIX, 373-390

Harris, M.; Raviv, A. (1972): Optimal Incentive Contracts with Imperfect Information, Journal of Economic Theory 20, 231-259

Holmström, B. (1979): Moral Hazard and Observability, The Bell Journal of Economics, 74-91

Kanodia, C. (1985): Stochastic Monitoring and Moral Hazard, Journal of Accounting Research, Vol. 23, No. 1, 175-193

LaValle, I.H. (1968): On Cash Equivalents and Information Evaluation in Decisions under Uncertainty, Part I,II,III, American Statistical Association Journal, 63, 252-290

Mattessich, R. (1984): Agency-Information Analysis and the New Science of Management Decisions, IFORS, Operations Research 84, 77-79

Mirrlees, J.A. (1976): The Optimal Structure of Incentives and Authory within an Organisation, The Bell Journal of Economics, 105-131

Penno, M. (1985): Informational Issues in the Financial Reporting Process, Journal of Accounting Research, Vol. 23, No. 1, 240-255

Ramakrishnan, R.T.S.; Thakor, A.V. (1982): Moral Hazard, Agency Costs, and Asset Prices in a Competitive Equilibrium, Journal of Financial and Quantitative Analysis, Vol. XVII, No. 4, 503-532

Shavell, S. (1979): Risk Sharing and Incentives in the Principal and Agent Relationship, The Bell Journal of Economics, 55-73

Simon, H. (1951): A Formal Theory of Employment Relation, Econometrica 19, 293-305

Singh, N. (1985): Monitoring and Hierarchies: The Marginal Value of Information in a Principal-Agent Model, Journal of Political Economy, Vol. 93, No. 3, 599-609.

Information Systems and the Design of Optimal Contracts

Marina Blickle

Summary: Contractual arrangements between principal and agent incorporate a number of signals which indicate the levels of output as well as various aspects of the agent's effort. The question is how many and which of the possible signals should be included in the contract which means that the salary paid to the agent will depend on the signals chosen. Firstly the paper focuses on costless signals. Secondly three cases are analysed where the costs of a signal vary with the degree of precision. In most cases the costs increase as the signals inform more about the agent's effort.

Table of contents:

1. Introduction

The economic theory of agency analyses contractual arrangements between principals and agents. It applies to any situation in which the outcome of the cooperation of principal and agent depends both on the random state of nature and on the action or effort chosen by the agent which is unknown to the principal (see Ross [1973]). Because of the informational asymmetry and the divergence of incentives between principal and agent, a departure from the optimal risk sharing solution is inevitable (see Rees [1985]).

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

The most prominent examples of principal-agent relations are those between owners and managers of a firm (see Jensen, Meckling [1976], Fama [1980]) and between employers and employees (see Stiglitz [1975]).

The managers' payment does not only consist of a fixed salary and some share of the collective output. In addition, some other criteria are considered in contracts offered to the managers by the owners of the firm. Such criteria are: observing certain ratios or indices, audited financial statements, the implementation of new technologies, and certain qualities of the managers (e.g. outward appearance, presentations, connections).

In many cases an employer can't observe his employees' performance. Most contracts between employer and employee explicitly stipulate only a fixed salary and sometimes some kind of profit-sharing. In addition, employees know that promotion and higher salaries often depend on further criteria. Examples of such criteria are: the demonstration of interest in research and innovations, participation in further education, reliability, personal relations to colleagues and superiors.

As we can see from these examples the agent's remuneration does not only consist of a fixed salary and some share of the output but is also made conditional on signals received from an information system. Information systems are established because they convey information about an unknown and unobservable capacity, attribute, or effort of the agent to the principal. The signals received from the information system are involved in the contract in order to discipline and motivate the agent to act in the principal's best interests. This kind of signal must be distinguished from signaling in the sense of Spence [1974] which may cause a separation of a pooled market.

The signals considered here may have properties quite different from signaling. In general they are connected with the action or effort and/or the random state of nature. In most cases information systems are not costlessly available. The costs vary with the degree of precision of the signal received from the information system.

The purpose of this paper is to analyse how the possibilities of acquiring information affect the structure of the optimal contract with special emphasis on the cost structure of the information system. The choice of the information system and the design of the contractual arrangement are interrelated and depend on setting off the benefits against the costs of the information system. Hence we have to answer the following questions:

(1) Is it favorable to involve a certain signal in the information system and how should the contract be made conditional on this signal?

(2) If different levels of precision of the signal are feasible which is the optimal level?

These questions will be analysed for a simple single-period setting. We will derive the optimal structure of a contract for the case of costless signals and analyse three cases of costly signals. Harris, Raviv [1979], Holmström [1979], Shavell [1979], Gjesdal [1982] and Singh [1985] have discussed the value of information systems, but only for the case of costless signals. In that case our model is a special case of the generalized agency model formulated by Gjesdal [1982].

2. The Model

We begin by setting out the model which will be used throughout the rest of the paper (see Spremann [1987]).

The agent chooses some action x which will be interpreted as effort. Together with the random state of nature θ unknown to the agent when x is chosen x determines the particular outcome Y.

In our model we assume that Y is given by the sum of x and θ,

$$Y = f(x, \theta) = x + \theta \quad . \tag{1}$$

An information system is composed of n possible signals $S_1, \ldots, S_n$. Each S_i may depend on x, θ and a random variable ε_i denoting the measurement error. For simplicity we define S_i by the sum of x and ε_i ,

$$S_i = g_i(x,\theta,\varepsilon_i) = x + \varepsilon_i \qquad i = 1,\ldots,n \,. \quad (2)$$

We further assume that θ and $\varepsilon_1,\ldots,\varepsilon_n$ are uncorrelated and normally distributed with mean 0 and variance σ_θ^2 and $\sigma_1^2,\ldots,\sigma_n^2$ respectively. σ_i^2 represents the precision of signal i ,

$$\begin{array}{lll} \theta \approx N(0,\sigma_\theta^2) & \varepsilon_i \approx N(0,\sigma_i^2) & \\ Cov(\varepsilon_i,\varepsilon_j) = \delta_{ij} & Cov(\varepsilon_i,\theta) = 0 & i,j = 1,\ldots,n \,. \end{array} \quad (3)$$

The payment schedule given by the contract can only depend on variables which both parties can observe, i.e. on Y and $S_1,\ldots,S_n$. For simplicity and because it is actually observed in many cases we assume a linear payment schedule ,

$$L = l(Y,S_1,\ldots,S_n) = c + dY + \sum_{i=1}^{n} e_i S_i \qquad c,d,e_1,\ldots,e_n \geqq 0 \,. \quad (4)$$

The costs of the information system (if there are any) depend on the level of precision of the signals ,

$$K_i = k_i(\sigma_i^2) \qquad i = 1,\ldots,n \,. \quad (5)$$

We suppose that x yields disutility to the agent; here the agent's opportunity costs are assumed to be a quadratic function of x ,

$$T = t(x) = x^2 \,. \quad (6)$$

Both principal and agent have von Neumann-Morgenstern utility functions u and v respectively. The principal is assumed to be risk-neutral; he maximizes his expected wealth. The agent is risk-averse with constant risk-aversion $\alpha > 0$, i.e. his utility function is $v(w) = -\exp(-\alpha w)$.

The principal is supposed to know all the characteristics of the agent and solves the following maximization problem:

(I) $\max E(Y-L)$

s.t.

(II) $E\{v(L-T-\sum_{i=1}^{n} K_i)\} \geqq m$

(III) $(x,\sigma_1^2,\ldots,\sigma_n^2) \in \operatorname{argmax} E\{v(L-T-\sum_{i=1}^{n} K_i)\}$.

A necessary condition for the agent to accept the contract is (II) because he insists on receiving at least some minimal expected utility m, the so-called reservation utility. The agent emits the signals, therefore he is the one who bears the information costs. However, these costs are compensated by the payment the agent receives from the principal. Since the choice of the payment schedule affects the agent's choice on effort, (III) must be satisfied if only incentive-compatible contracts are to be considered.

If we assume $k_1,\ldots,k_n$ to be differentiable, the first-order conditions yield the following properties of an inner solution:

(A) The marginal costs of the signal received from the information system are equal to the agent's marginal risk premium, i.e.

$$k_i'(\sigma_i^2) = -\frac{\alpha}{2} e_i^2 \qquad i = 1,\ldots,n .$$

(B) The agent's marginal opportunity costs are equal to his marginal expected payment. From this we can compute the optimal effort:

$$x = \frac{1}{2}(d + \sum_{i=1}^{n} e_i) .$$

(C) $(d,e_1,\ldots,e_n)$ is a solution of the following system of linear equations:

$$d(1 + 2\alpha\sigma_\theta^2) + \sum_{j=1}^{n} e_j = 1$$

$$d + \sum_{\substack{j=1\\ j\neq i}}^{n} e_j + e_i(1 + 2\alpha\sigma_i^2) = 1 \qquad i = 1,\ldots,n .$$

Since the agent's action or effort is not costlessly completely observable, costs accrue. In our model these so-called agency costs consist of

- the opportunity costs because of a socially non-optimal action or effort

- the risk premium required by the agent

- the costs of the information system (if there are any).

An optimal contract causes minimal agency costs. Hence in our model an optimal contract can be derived by solving the above noted maximization problem or by minimizing the agency costs.

3. Results

3.1 Costless Signals

We assume the information system to consist of n signals. (If a costless signal can be produced at several levels of precision, i.e. with different σ_i^2, we will only consider the highest level of precision, i.e. the smallest σ_i^2.)

In the optimal contract the coefficients $d, e_1, \ldots, e_n$ which denote the share of the output Y and the signals $S_1, \ldots, S_n$ respectively are positive and less than 1. Since the agent is risk-averse the sum of $d, e_1, \ldots, e_n$ is less than 1, too. Therefore the agent's optimal effort is less than in the case of costless observation and the agency costs are positive.

Result 1: The more precise the agent's effort is represented by Y or S_i (i.e. the smaller σ_θ^2 or σ_i^2) and the less the agent's risk-aversion the greater is the respective share and the smaller can the fixed salary c be chosen.

The more signals the information system contains the smaller are the respective shares. The shares remain positive and the sum of them grows as the number of signals increases. Hence we have a diversification effect: the optimal effort is increased and the agency costs are reduced.

Result 2: For $n \to \infty$ the agent's optimal effort approaches the value in the case of costless observation and the agency costs approach 0.

The principal should choose an information system which contains as many signals as possible. All of them are to be involved in the contract offered to the agent with their respective optimal share. Especially this is true if the agent is very risk-averse or if there are only very imprecise signals (i.e. signals with a great σ_i^2, i = 1,...,n). Otherwise we would have a socially very unsatisfying situation: The high fixed salary is independent of performance and the few incentives would lead to low effort by the agent and high agency costs.

3.2 Costly Signals

In the case of costly signals we cannot derive any general results. Depending on the parameters α, σ_θ^2 and above all on the costs of the signals (as a function of σ_i^2) it may even be optimal not to make the signal part of the information system and the payment schedule.

The different possible cases will now be demonstrated for the special case of only one costly signal.

Case 1: There are d discrete levels of precision s_j with costs $k(s_j) > 0$, $j = 1,\dots,d$.

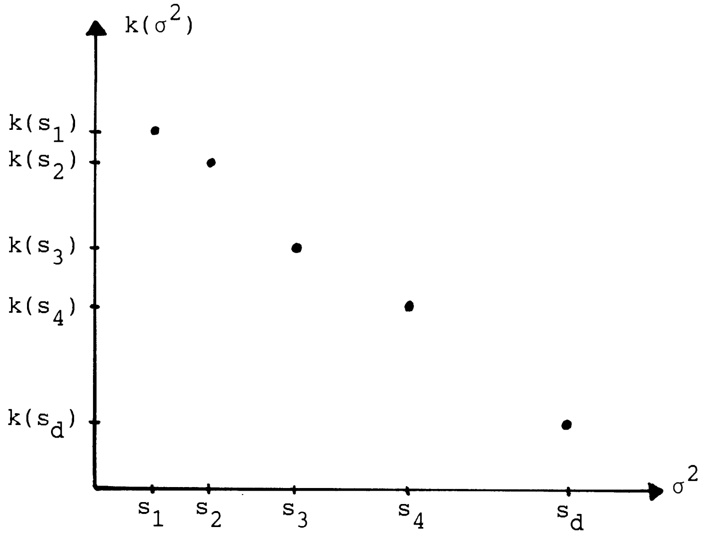


Figure 1: Discrete levels of precision (Case 1)

Depending on α, σ_θ^2 and $k(s_1),\ldots,k(s_d)$ it may be optimal (i.e. engendering minimal agency costs) either

(i) not to include the signal in the information system and the contract or

(ii) to make the signal part of the information system and the contract and the agent's optimal level of precision is s_l^*, $l \in \{1,\ldots,d\}$.

Case 2: Costs, linearly decreasing with a decreasing level of precision of the signal.

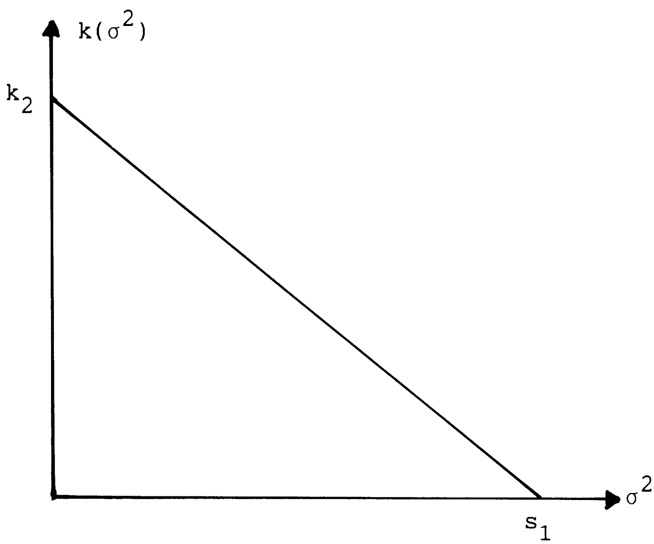


Figure 2: Linear costs of precision (Case 2)

There is no inner solution. Hence one of the extreme cases yields an optimal solution. The signal is included in the information system and made part of the payment schedule. Then it is optimal for the agent to signal either

(i) at level s_1 without costs (costless signals are always advantageous, see 3.1) or

(ii) his effort x exactly at cost $k_2 > 0$.

Case 3: The cost function is given by $k(\sigma^2) = \frac{1}{r\sigma^2}$, $r > 0$.

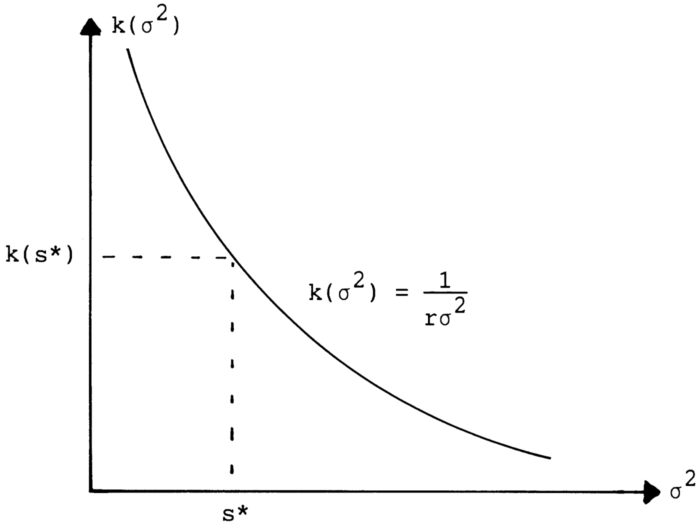


Figure 3: Inversely related precision costs (Case 3)

If

$$r \leqq \hat{r} = \frac{2(1 + 2\alpha\sigma_\theta^2)}{(\sigma_\theta^2)^2\alpha}$$

there is no inner solution. Hence it is optimal to do without this signal.

If $r > \hat{r}$ it is optimal to make the signal part of the information system and the contract. The agent will choose the level s^* with costs $k(s^*) > 0$ (given by the inner solution).

4. Conclusion

We have analysed contracts between principal and agent incorporating signals for a simple single-period setting in order to characterize the optimal structure of the information system and the payment schedule.

Our results indicate that costless signals are always advantageous. Hence the information system ought to comprise as many signals as possible, which are to be made part of the payment schedule with the respective optimal share. This generates a diversification effect (both an information and incentive effect) which causes an increase of the agent's effort and a reduction of the agency costs.

In the case of costly signals general statements are not possible. The design of the optimal information system and payment schedule depends on the agent's risk-aversion α, the variance of the random state of nature σ_θ^2 and the cost structure of the information system. Hence the following cases are possible:

(C1) A potential signal will not be included in the optimal information system and payment schedule.

(C2) The signal will be included and the agent signals

(C2a) at level $\sigma_i^{*2} > 0$ without cost;

(C2b) his effort x exactly to the principal at cost $k_i > 0$;

(C2c) at level $\sigma_i^{*2} > 0$ and at cost $k_i^* > 0$.

References

Fama, E.F. (1980): Agency Problems and the Theory of the Firm. Journal of Political Economy 88, 2, 288-307.

Gjesdal, F. (1982): Information and Incentives: The Agency Information Problem. Review of Economic Studies 49, 373-390.

Harris, M.; A. Raviv (1979): Optimal Incentive Contracts with Imperfect Information. Journal of Economic Theory 20, 231-259.

Holmström, B. (1979): Moral Hazard and Observability. The Bell Journal of Economics 10, 74-91.

Jensen, M.C.; W.H. Meckling (1976): Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. Journal of Financial Economics 3, 305-360.

Rees, R. (1985): The Theory of Principal and Agent Part I. Bulletin of Economic Research 37, 1, 3-26.

Ross, S.A. (1973): The Economic Theory of Agency: The Principal's Problem. American Economic Review 63, 2, 134-139.

Shavell, S. (1979): Risk Sharing and Incentives in the Principal and Agent Relationship. The Bell Journal of Economics 10, 55-73.

Singh, N. (1985): Monitoring and Hierarchies: The Marginal Value of Information in a Principal-Agent Model. Journal of Political Economy 93, 3, 599-609.

Spence, A.M. (1974): Market Signaling - Informational Transfer and Related Screening Processes. Harvard University Press, Cambridge (Mass.).

Spremann, K. (1987): Principal and Agent, in: G. Bamberg, K. Spremann (ed.), Agency Theory. Springer, Berlin, Heidelberg.

Stiglitz, J.E. (1975): Incentives, risk, and information: Notes toward a theory of hierarchy. The Bell Journal of Economics 6, 552-579.

Incentive Compatible Mechanisms for the Allocation of Public Goods

Helmut Funke

Summary: In the presence of public goods free market economies don't yield efficient allocations. For avoiding such a drawback several authors have proposed so-called incentive compatible mechanisms. The problem of finding such a mechanism coincides with the problem of a principal who is looking for a pay-off function that gives desired incentives to his agents. Unfortunately, the problem of incentive compatible mechanisms seems not to be solvable for economies with public goods.

Table of contents:

1. Introduction

One of the central issues of economic theory is the question how we can find an optimum of welfare. Although one could say that in a private ownership economy a competitive profit system would solve this problem in a rather good way this statement remains no longer true when public goods have to be taken into account. In this case we face the free rider problem resulting from the very property of public goods, i.e. the nonexcludability in consumption. In the presence of public goods free market economies often yield inefficient allocations, that is, there exists an (other) allocation where all members of the economy, agents say, are better off. Such an example is outlined in Section 2.

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

For achieving efficient states of economics with public goods many authors have proposed methods known as incentive compatible mechanisms. These are schemes of information and assessment where additionally a central coordination procedure is applied. The search for appropriate assessment functions (pay-off or tax functions) corresponds to the problem of a principal

- who does not know the agent's characteristics,
- who gets signals which are costless,
- where the agent's characteristics are independent of the signals.

Obviously this is a principal agent problem of a rather difficult kind.

Section 3 gives an introduction into tax schemes by presenting Lindahl's proposal. A formal definition of tax schemes is given in Section 4 where we will touch upon the very problems of such mechanisms. Section 5 presents a characterization of a certain class of mechanisms.

The discussion given in this paper shows a very pessimistic view of the incentive compatibility of mechanisms or tax schemes respectively. The suspicion that there does not exist an incentive compatible mechanism that really can work will be supported by this paper although it will not be proven in a strict sense.

2. Why Does a Competitive Profit System not Work?

For the illustration of this question just imagine a very simple economy with one private good, money say, and with one public good. All agents are supposed to have identical utility functions and equal initial endowments of private good. In addition let the initial quantity of the public good such that every individual's marginal utility equals the marginal cost of production. These assumptions imply that every agent feels at his/her utility maximum if he/she would have to pay alone for every additional unit of the public good. I.e. nothing would happen by private initiative. Unlike this result any case of cost sharing shows that some agents would be willing to finance additional quantity of the public good because their marginal cost now is (initially) lower than their marginal utility. This makes clear that in a competitive profit system the given initial state of economy may be stable but also that this state is inefficient because in the second case all agents would be better off than in the first case.

3. Lindahl's Tax Scheme

The earliest proposal for a tax scheme that takes individual preferences into account is that of Lindahl [1919] which is based on Wicksell's [1896] unanimity rule. It runs as follows:

(1) A central institution, government say, proposes individual tax rates $t_1,...,t_n$ that sum up to one.

(2) Agent i (= 1,...,n) faced to pay $t_i \cdot 100$ % of the costs of the public goods decides to vote for the quantity $y_i(t_i)$.

(3a) If unanimity holds, i.e. $y_1(t_1) = ... = y_n(t_n)$, then this plan will be performed.

(3b) If there are different votes for the quantity of public good, the government will lower the tax rates for those agents with the lowest votes and will raise the tax rates for those agents with the highest votes according to (1). Subsequently (2) will be performed again.

Provided that the agents are myoptic maximizers, i.e. $y_i(t_i)$ maximizes agents i's utility, we call a vector of tax rates fulfilling the unanimity rule (3a) a Lindahl equilibrium and the related allocation a Lindahl allocation. It can be shown that such an equilibrium belongs to the core of the economy. This implies

- Efficiency: There does not exist another allocation where no agent is worse off, and, at least one agent is better off than at the Lindahl allocation;

- Individual Rationality: All agents are not worse off than at the initial state.

Although it is usual to say that the payments are voluntary the Lindahl tax scheme gives no incentive "to tell the truth", i.e.

to tell the myoptic maximizers $y_i(t_i)$. In the long run, that is, if the procedure (1), (2), and (3b) converges only very slow, the agents will see their influence of their answers $y_i(t_i)$ upon the tax rates proposed the next time. Long sighted maximizing behavior requires a misrepresentation of the utilities, i.e. answers $\tilde{y}_i(t_i) < y_i(t_i)$. The viewpoint of a single agent i shows two consequences: First the final amount of the public good will decrease. Agent i would not be better of at his/her tax share concerning $y_i(\cdot)$. But second the tax share as well as the absolute tax decreases. An appropriate misrepresentation $\tilde{y}_i(\cdot)$ finally yields more utility than the proper $y_i(\cdot)$.

To come to the bad end, there exist economies where the myoptic maximizing behavior of only two agents means for the remaining n-2 agents the existence of dominant strategies $\tilde{y}_i(\cdot)$ yielding null tax rates for them.

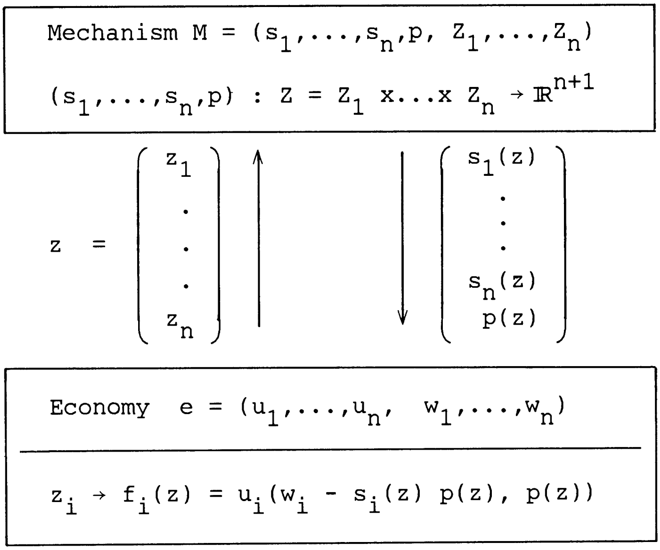
Strictly speaking such an unwanted excess is a consequence of the unanimity rule that gives the power of veto to each agent.

4. How are Incentive Compatible Mechanisms Constructed?

Obviously there is a need for a more efficient mechanism than Lindahl's approach. This fact has given in the 70es the impetus for the construction of several so-called incentive compatible mechanisms.

For the sake of simplicity we now deal with the case of only two commodities, that are one private and one public good. This simple case seemes not to be a severe restriction in view of the drawbacks that will be shown.

An incentive (compatible) mechanism can be represented by the following diagram



Z_i is the message space for agent i. Having collected all messages, i.e. $z = (z_1,...,z_n)$, the government calculates the individual tax rates $s_1(z),...,s_n(z)$, and the quantity $p(z)$ of the public good. As the mechanism $M = (Z_1,...,Z_n, s_1,...,s_n,p)$ is public, agent i can derive his/her utility

$$f_i(z) := u_i(w_i - s_i(z)\ p(z),\ p(z))$$

where u_i is his/her original utility function on the quantity space of private and of public good and w_i is the initial quantity of agent i's private good.

Obviously the combination of an economy $E = (u_1,...,u_n, w_1,...,w_n)$ with an mechanism $M = (Z_1,...,Z_n, s_1,...,s_n,p)$ yields a game theoretic situation, namely

$$G(f_1,...,f_n, Z_1,...,Z_n).$$

Here one is interested in the properties of <u>Nash equilibria</u>:

- does one exist?
- is it unique?
- are its strategies dominant?
- belongs its allocation to the core (efficiency at least)?
- is it stable?

Additionally there is a problem concerning the proper aim of such a mechanism: As government does not know the characteristics of the (present) economy, it seems necessary to require that a mechanism is all-rounded, i.e. that the resulting game shows all the nice properties listed above not only in the case of some few economies but for a set of economies as large as possible.

Unfortunately there is a tradeoff between the possibility of reaching efficient allocations and the strong incentive compatibility, i.e. where true utility functions or equivalent terms are dominant strategies in the message space. Following results of Hurwicz [1975], Groves [1979], and Walker [1980] there does not even exist a mechanism that for all economies with only separable and concave utility functions is strongly incentive compatible and that always yields efficient allocations. For example, as the well-known mechanism of Clarke [1971] and of Groves is strongly incentive compatible, it can no longer be surprising that it does not (necessary) balance the public budget, i.e. does not yield an efficient allocations.

Hurwicz', Groves', and Walker's results make clear that the decision for a mechanism that yields efficient allocations means that it only can be weakly incentive compatible, that is, there will exist at best only an ordinary Nash equilibrium of which the allocation is efficient. Indeed this modest aim is attained by the mechanisms of Groves and Ledyard [1977], Hurwicz [1979], and Walker [1981]. But if "telling the truth" (in some sense) is not a dominant strategy the question about strategic behavior becomes important again. This yields the requirement of stability as pointed out in the following: If dominant strategies exist it makes sense to choose them no matter what other agents are doing. Therefore we have no stability problem in the case that these dominant strategies are known by the agents, e.g. see the Clarke-Groves mechanism. But, if there are no dominant strategies we have the problem to find the Nash equilibrium by a search process. If convergence is to slow then the agents may behave in a strategic manner by recognizing the reactions of the other agents. For avoiding such behavior the speed of convergence should be as high as possible. Unfortunately the weak incentive compatible mechanisms given by Groves-Ledyard, Hurwicz, and Walker are unworkable just because of the stability question. The Groves-Ledyard mechanism can yield arbitrary many Nash equlibria, see Bergstrom, Simon, and Titus [1983], and it converges only under conditions that are not acceptable, see Muench and Walker [1983]. The mechanisms of Hurwicz and Walker are not stable at all. In the case of Walker's mechanism Nash adjustment strategies yield difference equations where a local analysis shows that the concerning functional matrix has the eigenvalue 1-n.

5. Characterization of a Certain Class of Tax Schemes

As we have seen in the previous section we should doubt whether there exist mechanisms that are weakly incentive compatible as well as stable. As we will point out the answer will be "no" for a set of reasonable properties. Although this result is not proven for all economies it strengthens our suspicion that there is no way to solve the free rider problem for a set of egoistic agents in the presence of public goods, i.e. in general allocations of the core are not attainable.

We consider the following class of mechanisms:

$$(P0)\quad \begin{cases} M = (s_1,\dots,s_n,p) \; : \; Z = Z_1 \times \dots \times Z_n \to \mathbb{R}^{n+1}_{++} \\ Z_i = \mathbb{R} \text{ for } i = 1,\dots,n, \\ M = \text{ is differentiable once,} \\ \frac{\partial p}{\partial z_i}(z) > 0 \quad \text{for all } z \in Z \text{ and } i = 1,\dots,n\,. \end{cases}$$

Additionally we require the following properties:

(P1) Weak Progression of Tax:

$$\frac{\partial s_i}{\partial z_i}(z) \geqq 0 \quad \text{for all } z \in Z \text{ and } i = 1,\dots,n\,.$$

(P2) Feasilibity: All Nash equilibria $\bar{z}$ of the related games, G say (see section 4), yield feasible allocations:

$$\sum_{i=1}^{n} s_i(\bar{z}) \geqq 1\,.$$

(P3) Efficiency: All Nash equilibria yield Pareto optimal allocations.

Property (P0) means among others that every agent has only one vote for the quantity of public good. Property (P1) in connection with (P0) means that an agent has to face tax rates as higher as his/her desire for public good is. Property (P2) requires financing without deficit, and property (P3) is a standard of economic theory that speaks for itself.

As will be shown in Theorem 1 these properties are equivalent to the following two:

(P4) Balanced Budget:

$$\sum_{i=1}^{m} s_i(z) = 1 \quad \text{for all } z \in Z .$$

(P5) No Direct Influence to Tax Rates:

$$\frac{\partial s_i}{\partial z_i}(z) = 0 \quad \text{for all } z \in Z \quad \text{and } i = 1,\dots,n .$$

Property (P4) says that there is never an excess tax yield. Property (P5) means that the message of agent "i" has no influence on his/her own tax rate. For proving the following theorem we utilize economies with separable utility functions:

(E) The utility functions have the following form:

$$u_i(x_i y) = x_i + \phi_i(y) \quad \text{for } i = 1,\dots,n ,$$

where ϕ_i is strictly concave and monotonic increasing.

Theorem 1: A mechanism of the type (P0) meets the properties (P1), (P2), and (P3) for all economies of the type (E) iff it meets the properties (P4) and (P5) for all economies of (E).

<u>Sketch of the proof:</u> Given an arbitrary but fixed $\bar{z} \in Z$ because of (P0) and (P1) one can construct valuation functions $\phi_1,\ldots,\phi_n$ such that $\bar{z} = (\bar{z}_1,\ldots,\bar{z}_n)$ is the unique Nash equilibrium. The construction of ϕ_1 (i = 1, w.l.o.g.) runs as follows: Let c_1 the inverse of $p(\cdot,\bar{z}_{-1})$ where $\bar{z}_{-1}$ is given by $(\bar{z}_2,\ldots,\bar{z}_n)$. So $y = \bar{y}$ with $c_1(\bar{y}) = \bar{z}_1$ maximizes

$$f_1(c_1(y),\bar{z}_{-1}) = w_1 - s_1(c_1(y),\bar{z}_{-1})y + \phi_1(y) .$$

For example, this will be fulfilled for ϕ_1 given by

$$\phi_1'(y) := e^{\bar{y}-y}\, s_1(\bar{z}) + \begin{cases} \max\limits_{x\in[y,\bar{y}]} \dfrac{\partial s_1}{\partial z_1}\Big(c_1(x),\bar{z}_{-1}\Big)\, c_1'(x)x & \text{for } y < \bar{y} \\[2ex] \min\limits_{x\in[\bar{y},y]} \dfrac{\partial s_1}{\partial z_1}\Big(c_1(x),\bar{z}_{-1}\Big)\, c_1'(x)x & \text{for } y \geqq \bar{y} \end{cases} .$$

It easily can be seen that ϕ_1' is positive valued and strictly monotonic decreasing. Therefore ϕ_1 is strictly monotone increasing and strictly concave. Furthermore, for given $\bar{z} = (\bar{z}_1,\ldots,\bar{z}_n)$ the construction of ϕ_1 does not depend on $\phi_2,\ldots,\phi_n$, i.e., we can construct the valuation functions independently to each other.

For proving property (P4) we take an arbitrary $z \in Z$ and a suitable economy where, as just shown z is the unique Nash equilibrium of the related game. Obviously, efficiency does not allow for a ">" in the condition of feasibility, (P2). Therefore "=" holds for all $z \in Z$, i.e. (P4).

For showing property (P5) again let an arbitrary $z \in Z$ be the unique Nash equilibirum of the related game of a suitable economy. The ordinary set of necessary conditions is

$$-\frac{\partial s_i}{\partial z_i}(z) \cdot \frac{p(z)}{\frac{\partial p}{\partial z_i}(z)} - s_i(z) + \phi_i'(p(z)) = 0 \quad \text{for } i = 1,\ldots,n .$$

Because of (P4) and because of Samuelson's [1954] condition of efficiency, that is

$$\sum_{i=1}^{n} \phi_i'(y) = 1,$$

we obtain

$$\sum_{i=1}^{n} \frac{\partial s_i}{\partial z_i}(z) \cdot \frac{\partial p(z)}{\frac{\partial p}{\partial z_i}(z)} = 0 \quad .$$

As every term in this sum has the same sign because of (P0) and (P1)

$$\frac{\partial s_i}{\partial z_i}(z) = 0$$

holds. Now we have shown that (P0), (P1), (P2), and (P3) imply (P4) and (P5).

The converse is also true, i.e. (P0), (P4), and (P5) imply (P1) and (P2). Because of (P4) and (P5) it easily can be shown that Nash equilibria yield Lindahl allocations. As every Lindahl equilibrium belongs to the core of the economy efficiency, (P3), is clearly given. Finally, (P1) and (P2) are trivially consequences of (P4) and (P5). ∎

As just pointed out at the end of the previous proof mechanisms with the properties (P0), (P1), and (P3) yield Lindahl allocations. On the one hand side this result seems to be a very desirable one because of its implications to the welfare of the economy, on the other hand side this result is not at all desirable because the possibilities and the successes of manipulation:

As it is known from now on that Lindahl allocations will come out a tricky agent has nearly the same possibilities for misrepresenting his/her utility no matter if Lindahl's proposal is performed or a mechanism given by Theorem 1. Such an egoistic agent simply has to transform his/her reactions from one mechanism to the other, then the result will be the same for him/her.

There is to be only a <u>slight difference</u>: By Lindahl's proposal every agent has the power of veto. But given a mechanism in the sense of Theorem 1 no one can do better than react in the myoptic way provided that the other agents choose their Nash equilibrium strategies in an unbiased way.

This statement makes the requirement for stability indespensable, i.e. Nash equilibria should be reached as quickly as possible. Unfortunately, convergence never is the case:

Theorem 2: Every mechanism with the properties (P0), (P4), and (P5) is unstable for symmetric economies that additionally meet property (E).

Sketch of the proof: Let

$$u_i(x_i,y) = x_i + \phi(y)$$

the utility functions and let $\bar{z}$ the (unique) Nash equilibrium. Nash adaption means that z_i^{t+1} maximizes

$$w_i - s_i(z^t)\, p(z_1^t,\ldots,z_{i-1}^t,\cdot,z_{i+1}^t,\ldots,z_n^t) +$$

$$+ \phi(p(z_1^t,\ldots,z_{i-1}^t,\cdot,z_{i+1}^t,\ldots,z_n^t)) \quad .$$

A local analysis of this difference equation in a sufficiently small neighborhood of $\bar{z}$ shows that

$$\left(\frac{\partial p}{\partial z_1}(\bar{z}),\ldots,\frac{\partial p}{\partial z_n}(\bar{z})\right)$$

is a left eigenvector with the eigenvalue 1-n of the Jacobian of the reaction functions resulting from Nash adaption, i.e. it is not at all stable.

References:

Bergstrom, T., C.P. Simon and C.J. Titus [1983]. Counting Groves-Ledyard Equilibria Via Degree Theory. Journal of Mathematical Economics 12, 167-184.

Clarke, E.H. [1971]. Multipart Pricing of Public Goods. Public Choice 11, 17-33.

Groves, T. and J. Ledyard [1977] . Optimal Allocation of Public Goods: A solution to the 'Free Rider Problem'. Econometrica 45, 783-809.

Groves, T. [1979]. Efficient Collective Choice when Compensation is Possible. Review of Economic Studies 46, 227-241.

Hurwicz, L. [1975]. On the Existence of Allocation Systems Whose Manipulative Nash Equilibria are Pareto-Optimal. Unpublished Paper.

Hurwicz, L. [1979]. Outcome Functions Yielding Walrasian and Lindahl Allocations at Nash Equilibrium Points. Review of Economic Studies 46, 217-225.

Lindahl, E. [1919]. Die Gerechtigkeit der Besteuerung. Lund.

Muench, R. and M. Walker [1983]. Are Groves-Ledyard Equilibria Attainable. Review of Economic Studies 50, 393-396.

Samuelson, P. [1954]. The Pure Theory of Public Expenditure. Review of Economics and Statistics 36, 387-389.

Walker, M. [1980]. On The Nonexistence Of A Dominant Strategy Mechanism For Making Optimal Public Decisions. Econometrica 48, 1521-1540.

Walker, M. [1981]. A Simple Incentive Compatible Scheme For Attaining Lindahl Allocations. Econometrica 49, 65-71.

Wicksell, K. [1896]. Finanztheoretische Untersuchungen. Jena.

Incentives to Forecast Honestly

Andreas Pfingsten

Summary: Situations are shown where an informational asymmetry prevails in a principal-agent relation. A general model is formulated to derive some first results in this framework. Modified assumptions yield a similar but more simpler model and further insights into the problem. Yet the reader is left with a considerable research agenda.

Table of Contents:

1. Introduction

In the literature on the principal-agent relation it is predominantly assumed that principal and agent have identical beliefs about the probabilities of the occurrence of possible states of nature. Rees [1985a, p.5] mentions in his survey that this assumption, of course, is a major limitation, even if it should be justified in many real-world cases. The recent contribution by Beckmann [1987] shares our view that instead there is an information asymmetry, e.g., in the owner-manager relation.

In the present paper we will look at a situation where the beliefs do not coincide. Information will be distributed asymmetrically: the agent knows the true density function of the random variable, the principal does not have any knowledge about it at all (except that its values must all be from a certain range).

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

First, we will briefly describe a few situations where this asymmetry seems to be quite plausible (section 2). In section 3 a general model will then be introduced formally. Some results are derived in section 4. Simplifying assumptions of the general model, still pertaining the information asymmetry, will yield to a case which is much easier to handle and will give rise to further insights (section 5). A research agenda concludes.

This paper is written with a very specific objective in mind which is different from the usual proceedings volume motivation. It is intended to deal with some information asymmetry as a relatively new feature of principal-agent models and to show some of the implications and problems arising. Although much of the analysis is done quite formally, the reasoning in the paper several times is very informal. We have deliberately chosen mainly to share the flavour of this topic with all our readers and hence leave rigorous proofs of results derived to subsequent work. Scholars will hopefully be happy that we present many interesting questions, and even more happy since we leave many of them unsolved.

2. Some Applications

Suppose you are the planner in some centrally planned economy. Then you are interested to know how many units of output the single firms can produce. You need this data in order to come up with feasible production and consumption plans for the whole economy. In the current socialist countries, however, you will find the unfortunate situation that plant managers are not willing to inform the planner honestly about their output expectation. It is neither to their nor to their workers' benefit if they reveal the true working potential, since the values they report as expected output are, after some modifications by the central planning agency, used as a measuring stick for the firm's performance. Hence lower values mean that less work effort is necessary to meet the government's requirements, and it is easier to achieve honourable mention or financial rewards. (For material on incentive systems in socialist economies see, e.g., Gerhardt [1967], Weitzman [1976], Rees [1985b].)

As a second example suppose you are in charge of coordinating the production of a multinational firm with plants in different countries. If the plant managers are rated according to how they have met their output target, then the situation is basically the same as in the case described above.

Similarly, asking salesmen to predict their expected sales will, given they know that their reported expectation will be used to determine sales requirements, also tend to produce underestimations of the true expected sales. If each salesman is working in a separate region then, equivalently, he will try to picture the market potential of the product in his region as bad as possible. The excess of actual sales over planned sales could then be attributed to chance and extra effort rather than to his cheating.

Finally, suppose a firm grants discounts to stores where its products are sold, the discounts are initially calculated on the basis of total expected sales (discounts increasing in sales), and be the expectation just the value reported by the store at the beginning of the period. If unjustified discounts (sales expectation not actually realized) have simply to be repaid at the end of the period without an extra fine then there is a strong incentive for the store to exaggerate his sales ability.

All of the cases mentioned have common properties. There are two parties acting, one is better informed than the other. The less informed party (henceforth the principal) is paying the more informed party (the agent), and the pay, respectively the reward, discount etc., depends at least partly on a value which the more informed party reports. Since there is no incentive to reveal true knowledge the agent will (in his own interest) not say the truth.

3. The Model

Agent and principal both know that the output x lies in an interval $I=[m,M]$, where $0 < m < M < \infty$. Output is a random variable, and the agent, but only he, knows the continuous density function $f: I \to \mathbb{R}_+$.

The agent has a utility function $V:\mathbb{R}_+\rightarrow \mathbb{R}$. The principal only knows about the agent's utility function that it is a strictly increasing, strictly concave function of income and that the agent's objective is to maximize expected utility.

The principal has a strictly increasing, (weakly) concave utility function $U:\mathbb{R}_+\rightarrow \mathbb{R}$ depending on his income. He tries to maximize expected utility as well.

At the beginning of the planning period the agent reports which output a he expects for this period. He does not know the state of the nature (and hence the actual output x) by that time. His income is determined as y(x,a), where $y:I\times I\rightarrow \mathbb{R}_+$ is the payment function given by the principal.

The agent's control variable consequently is his reported output expectation a, whereas the principal tries to optimize by choice of the function y. The latter has to take into account when deciding on y that he can sell every unit of output produced for a given market price $p>0$ and that he will have to pay, apart from the agent's income, production and adjustment costs c(x,a). The principal knows the cost function $c:I\times I\rightarrow \mathbb{R}_+$ when choosing the function y, for the agent's decision, however, this function is irrelevant. To this end, the functions c and y will be assumed to be sufficiently (continuously) differentiable.

(Often it is assumed that output (and hence income) depends on the agent's effort, where it is furthermore suspected that more effort means less utility. It is well possible to enrich the present analysis by incorporating such assumptions, e.g.: The density function f is replaced by density functions f^e representing, for different e, different levels of effort. Then one can determine an optimal value a^e for each f^e. Among these values a^e a value a^* must be chosen using a modified utility function. Instead of applying such an approach it is also conceivable to simply complement the list of variables of the functions f and V by an effort variable e.)

For a given payment function y the expected utilities of principal and agent, respectively, depend on the agent's reported value a and are given by the following expressions:

$$P(a) = \int_m^M U[p \cdot x - y(x,a) - c(x,a)] \cdot f(x)\, dx \, . \qquad (3.1)$$

$$A(a) = \int_m^M V[y(x,a)] \cdot f(x)\, dx \, . \qquad (3.2)$$

4. Some Results

If the agent maximizes his expected utility by choice of a, then a necessary condition for an interior solution a^* is

$$A'(a^*) = \int_m^M V'[y(x,a^*)] \cdot y_a(x,a^*) \cdot f(x)\, dx = 0 \, , \qquad (4.1)$$

where y_a denotes the first-order partial derivative of y with respect to a. For a (local) maximum it is sufficient if in addition

$$A''(a^*) = \int_m^M [V''[y(x,a^*)] \cdot [y_a(x,a^*)]^2 + V'[y(x,a^*)] \cdot y_{aa}(x,a^*)] \cdot f(x)\, dx < 0 \qquad (4.2)$$

holds, where y_{aa} denotes the second-order partial derivative of y with respect to a. In case of

$$A''(a) < 0 \qquad \text{for all } a \in I \qquad (4.3)$$

a value a^* determined from condition (4.1) yields the unique (global) maximum. We have

Fact 1

If the payment function y is strictly concave in the reported value a for all output levels x, then a value a^* satisfying condition (4.1) is the (unique) best value for the agent to report.

For a proof of this assertion one should note that $V''(a) < 0$ for all a, $V'(a) > 0$ for all a, and $f(x) \geq 0$ for all x. Similarly, we see

<u>Fact 2</u>

If the payment function y is strictly convex in the reported value a for all output levels x, then there exist utility functions V such that a value a^* satisfying condition (4.1) yields a (local) minimum of the function A.

In order to achieve the result reported in Fact 2, a function V needs to be "sufficiently little" concave.

If the principal wants that for the agent always, i.e. for arbitrary density functions f, a unique global maximum of expected utility exists, then he must choose a payment function y which is for all output levels x strictly concave in the reported value a. This is the case since, as kind of a strengthening of Fact 2, we have

<u>Fact 3</u>

If the payment function y is strictly convex in reported values a for output levels x from an interval $\tilde{I} \subseteq I$ then there exist density functions f and utility functions V such that a reported value a^* satisfying condition (4.1) yields a local minimum of the function A.

Such density functions may have, e.g., the property f(x) = 0 for all $x \notin \tilde{I}$.

If the payment function y does not depend on a, e.g.,

$$y_a(x,a) = 0 \qquad \text{for all } x,a \in I \ , \tag{4.4}$$

then condition (4.1) trivially holds, but condition (4.2) is violated. In this case it does not matter for the agent's income which value he reports. Thus, he does not have any incentive at all to reveal his true expectation.

If the payment function y is strictly concave in the reported value a and condition (4.1) is satisfied then it must be true for a^* that

$$\exists\ \delta > 0: \quad y_a(x,a^*) > 0 \quad \text{for all } x < \delta \tag{4.5}$$

since otherwise strict concavity would imply for all $x \in I$

$$y_a(x,a^*) \leqq 0 \qquad \text{with strict inequality for all } x > m. \tag{4.6}$$

But then $A'(a^*)$ was negative. If the density function has the property

$$f(x) = 0 \qquad \text{for all } x > \tau \leqq \delta \tag{4.7}$$

and if, in addition, condition (4.5) holds then again a^* can be no solution since $A'(a^*)$ this time was positive. Consequently, there also must be values x such that

$$y_a(x,a^*) < 0 \quad \text{and} \quad f(x) \neq 0\ . \tag{4.8}$$

It is immediately plausible that the payment, given some arbitrary fixed output level x, must first increase in the reported value a and then decrease. By strict concavity we even know that the marginal increase must be decreasing.

Let us now look at the utility maximization problem of the principal, and let us assume that there exists an optimal payment function y^* which the principal has found. (As a matter of fact, there may not exist such a y^*, and it is an important question for further research to determine necessary and sufficient conditions for the existence. It should be examined, e.g., what the consequences are if the agent must be guaranteed a minimum standard of living.) Given y^*, the principal's expected utility depends on the agent's reported value a and is given by expression (3.1).

It now would be in the principal's interest if, given y^*, the agent's reported value a^* determined according condition (4.1) would maximize P(a) as well. Similar to conditions (4.1) and (4.2) for the agent, we have sufficient conditions for a maximizer a^P of the principal's expected utility:

$$P'(a^P) = \int_m^M U'[p\cdot x - y^*(x,a^P) - c(x,a^P)] \cdot [- y_a^*(x,a^P) - c_a(x,a^P)] \cdot f(x)\, dx$$
$$= - \int_m^M U'[p\cdot x - y^*(x,a^P) - c(x,a^P)] \cdot [y_a^*(x,a^P) + c_a(x,a^P)] \cdot f(x)\, dx \quad = 0, \tag{4.9}$$

$$P''(a^P) = - \int_m^M [U''[p\cdot x - y^*(x,a^P) - c(x,a^P)] \cdot [y_a^*(x,a^P) + c_a(x,a^P)]^2 \cdot (-1) + U'[p\cdot x - y^*(x,a^P) - c(x,a^P)] \cdot [y_{aa}^*(x,a^P) + c_{aa}(x,a^P)]] \cdot f(x)\, dx \quad < 0 . \tag{4.10}$$

It is immediately seen that each value a^P satisfies condition (4.9) that satisfies the sufficient condition

$$y_a^*(x,a^P) + c_a(x,a^P) = 0 \qquad \text{for all } x\in I . \tag{4.11}$$

If the principal wants to choose a payment function y^* such that each a^* which is maximizing the agent's expected utility is also utility maximizing for himself then an agent's optimal reported value a^* must always satisfy condition (4.9). The principal can be sure that this is the case if each such a^* satisfies condition (4.11), i.e.,

$$y_a^*(x,a) + c_a(x,a) = 0 \qquad \text{for all } a,x\in I . \tag{4.12}$$

By integration with respect to the variable a, this first-order differential equation yields

$$y^*(x,a) \equiv - c(x,a) + d(x) . \tag{4.13}$$

This result is very important economically. It means that the principal levies all adjustment costs on the agent. Hence the agent bears all the risk from misprediction of output. And it is therefore obvious intuitively as well that it is not in the agent's interest to report something different from his true expectation. The principal's expected utility in this situation is independent of the agent's reported value a^*, namely

$$P(a^*) = \int_m^M U[p \cdot x - [- c(x,a^*) + d(x)] - c(x,a^*)] \cdot f(x)\, dx = \int_m^M U[p \cdot x - d(x)] \cdot f(x)\, dx \,. \tag{4.14}$$

The agent's expected utility is

$$A(a^*) = \int_m^M V[d(x) - c(x,a^*)] \cdot f(x)\, dx \,. \tag{4.15}$$

One question is in which other cases an (interior) solution of the principal's problem, yielding a global maximum of P given y^*, exists. Another question (we are not going to deal with for the moment) is whether, respectively when, such a solution a^P of the principal's problem coincides with the solution a^* of the agent's problem.

Looking at the signs of the expressions in condition (4.10) we find:

<u>Fact 4</u>

If the adjustment cost function c is independent of the reported value a and if the optimal payment function y^* is strictly convex in a for all x then condition (4.9) implicitly determines the agent's reported value a^P which is the unique global maximum for the principal.

If the payment function is strictly concave in a then it is not guaranteed that the principal's expected utility is maximized by the agent. The economic importance of this result is striking: If adjustment costs do not depend on the agent's reported value a, i.e., if there are no adjustment costs, then only a payment function which is strictly <u>convex</u> in a guarantees optimality of a solution derived from condition (4.9), whereas optimality of a solution a^* derived from condition (4.1) is only guaranteed if the payment function is strictly <u>concave</u>. In other words:

<u>Fact 5</u>

If the adjustment cost function c is independent of the agent's reported value a then there does not exist any payment function which guarantees that a solution a^* derived from condition (4.1) as well as a solution a^P derived from condition (4.9) are both optimal.

The analysis of the principal's problem can be performed in different notation to yield some further insights. Define

$$K^*(x,a) \equiv y^*(x,a) + c^*(x,a) \ . \tag{4.16}$$

Then we have:

$$P(a) = \int_m^M U[p \cdot x - K^*(x,a)] \cdot f(x)\, dx \ , \tag{4.17}$$

$$P'(a) = - \int_m^M U'[p \cdot x - K^*(x,a)] \cdot K_a^*(x,a) \cdot f(x)\, dx = 0 \ . \tag{4.18}$$

$$\begin{aligned} P''(a) &= - \int_m^M [U''[p \cdot x - K^*(x,a)] \cdot [K_a^*(x,a)]^2 \cdot (-1) \\ &\qquad + U'[p \cdot x - K^*(x,a)] \cdot K_{aa}^*(x,a)] \cdot f(x)\, dx \\ &= \int_m^M [U''[p \cdot x - K^*(x,a)] \cdot [K_a^*(x,a)]^2 \\ &\qquad - U'[p \cdot x - K^*(x,a)] \cdot K_{aa}^*(x,a)] \cdot f(x)\, dx < 0 \ . \end{aligned} \tag{4.19}$$

This notation is quite useful, for example when analyzing further sufficient conditions for solutions of the principal's problem. In addition it is seen how closely the principal's conditions (4.18) and (4.19) are related to the agent's conditions (4.1) and (4. 2). In the agent's problem concavity of y was important while now convexity of K matters. (Of course, the controls of agent and principal, reported value a and payment function y, respectively, are very different.)

In the adjustment cost literature (see Brechling [1975, chapter5] for an example) it is often assumed that functions like c are convex. A payment function y^* which is "sufficiently little" concave would then be capable of guaranteeing convexity of K^*, e.g.,

$$y^*(x,a) \equiv \lambda \cdot c(x,a) \qquad (0 < \lambda < 1) \ . \tag{4.20}$$

An interesting special case is that of a risk-neutral principal (U'' vanishes). For an interior maximum it then indeed is important that the cost function K^* is strictly convex. Furthermore, looking at condition (4.19) one can suspect that increasing risk aversion of the principal (locally) allows the payment function to be "less" convex without harming existence of an interior maximum for the principal.

We conclude this section by pointing to two problems that remain: First, so far we have not required that the agent always receives an income above some given (e.g., poverty) level. This is certainly necessary for practical applications. Second, we have not looked at the choice of the optimal payment function at all, but simply have assumed their (unique) existence. Yet it is conceivable that there are many payment functions y such that the agent's optimal reported value a, given the payment function, also maximizes the principal's expected utility. It may also be the case that there does not exist any such function. In all these cases the principal must perform kind of a second-best analysis.

5. A Simple Case

In this section we are going to examine a simple case of the general model. It is based on assumptions which are partly different from those in the earlier parts of the paper.

To exclude questions concerning risk-sharing, agent and principal are both assumed to be risk-neutral. The principal chooses the payment function

$$y(x,a) = w \cdot x - \begin{cases} s^1 \cdot (x-a) & \text{for } a \leq x \quad (s^1>0) \\ s^2 \cdot (a-x) & \text{for } a > x \quad (s^2>0) \end{cases} \quad (w > 0). \qquad (5.1)$$

The agent hence will try to maximize, by choice of the reported value a,

$$A(a) = \int_m^a [w \cdot x - s^2 \cdot (a-x)] \cdot f(x)\,dx + \int_a^M [w \cdot x - s^1 \cdot (x-a)] \cdot f(x)\,dx. \qquad (5.2)$$

From

$$A(a) = \int_m^M w \cdot x \cdot f(x)\,dx + \int_m^a s^2 \cdot x \cdot f(x)\,dx - \int_a^M s^1 \cdot x \cdot f(x)\,dx - \int_m^a s^2 \cdot a \cdot f(x)\,dx + \int_a^M s^1 \cdot a \cdot f(x)\,dx \qquad (5.3)$$

the following necessary condition for an interior solution a^* is derived:

$$A'(a^*) = s^2 \cdot a^* \cdot f(a^*) + s^1 \cdot a^* \cdot f(a^*) - \int_m^{a^*} s^2 \cdot f(x)\,dx - a^* \cdot s^2 \cdot f(a^*) + \int_{a^*}^M s^1 \cdot f(x)\,dx - a^* \cdot s^1 \cdot f(a^*) = 0 . \qquad (5.4)$$

Since this easily simplifies to the condition

$$- s^2 \cdot \int_m^{a^*} f(x)\,dx + s^1 \cdot \int_{a^*}^M f(x)\,dx = 0 \qquad (5.5)$$

the central criterion

$$F(a^*) = \frac{s^1}{s^1 + s^2} \qquad (5.6)$$

is derived, where F is the distribution function corresponding to the density function f. Such a condition is known from the statistical decision theory (e.g., Bamberg/Baur [1980, p.254]). It is obvious that with an increase of s^1 (s^2), the other parameter remaining constant, the solution a^* must weakly increase (decrease). Economically this means that it is the principal's choice whether overestimation or underestimation of the true result is encouraged.

If the parameters s^1 and s^2 are less than w the payment function has the important property to preserve the agent's motivation to work: If some value a has been reported it is still beneficial for the agent if the actual output x is as big as possible. Even if the income per unit of output decreases once x is greater than a, total income still increases in output.

It should be noted that in case of $f(a^*) \neq 0$ the solution is unique and because of

$$\begin{aligned} A''(a^*) &= - s^2 \cdot f(a^*) - s^1 \cdot f(a^*) \\ &= - (s^1 + s^2) \cdot f(a^*) \end{aligned} \tag{5.7}$$

indeed yields the maximum. If the density function f is sufficiently differentiable then, because of

$$f(x) \geqq 0 \quad \text{for all } x \in I , \tag{5.8}$$

we must have $f'(a^*)=0$. For f'' the same reasons imply that only the following two cases are possible:

$$f''(a^*) > 0 , \tag{5.9}$$

$$f''(a^*) = 0 . \tag{5.10}$$

In the first case we have a unique maximum as well. In the second case there might exist a neighbourhood around a^* where we find values $\tilde{a}$ yielding the same expected income for the agent as a^* does. In this case, of course, the value a^* determined by condition (5.6) is not unique.

The principal's expected income is

$$\begin{aligned} P(a) &= \int_m^a [p \cdot x - c(x,a) - w \cdot x + s^2 \cdot (a-x)] \cdot f(x) \, dx \\ &\quad + \int_a^M [p \cdot x - c(x,a) - w \cdot x + s^1 \cdot (x-a)] \cdot f(x) \, dx \\ &= (p - w) \cdot E(x) - \int_m^M c(x,a) \cdot f(x) \, dx \\ &\quad + s^2 \cdot \int_m^a (a-x) \cdot f(x) \, dx + s^1 \cdot \int_a^M (x-a) \cdot f(x) \, dx . \end{aligned} \tag{5.11}$$

It would be in the principal's interest if the agent reported a value a^P satisfying

$$P'(a^P) = -\int_m^M c_a(x,a^P) \cdot f(x)\,dx \tag{5.12}$$

$$+ s^2 \cdot \left[\int_m^{a^P} f(x)\,dx + a^P \cdot f(a^P) - a^P \cdot f(a^P) \right]$$

$$+ s^1 \cdot \left[- a^P \cdot f(a^P) - \int_{a^P}^M f(x)\,dx - a^P \cdot (-f(a^P)) \right]$$

$$= -\int_m^M c_a(x,a^P) \cdot f(x)\,dx + (s^1 + s^2) \cdot F(a^P) - s^1 = 0 \; .$$

If the agent reports a value a^* maximizing his own expected income then the principal's expected income will be minimized, e.g., when the cost function c does not depend on a. The principal's expected income will be maximized, e.g., when the cost function c is "sufficiently convex" in a and in addition satisfies, for the density function f,

$$\int_m^M c_a(x,a^*) \cdot f(x)\,dx = 0 \; . \tag{5.13}$$

We do not know so far whether there exist payment functions y different from (5.1) that are even better for the principal. And one must observe that while the principal's expected income is increasing in s^1 and s^2 the motivation problem mentioned earlier imposes an upper bound that he might not want to exceed.

Suppose the cost function is of the type

$$c(x,a) = (a - x)^2 \tag{5.14}$$

which is often applied. Condition (5.13) then becomes

$$\int_m^M 2 \cdot (a^* - x) \cdot f(x)\,dx = 0 \tag{5.15}$$

and eventually

$$a^* = E(x) \; . \tag{5.16}$$

Economically this means that the principal would like to have the agent reveal his (presumably correct) output expectation. If the true density function is symmetric, an assumption which in many relevant cases probably is not too bad, then the principal must simply choose $s^1 = s^2$.

The second-order condition for the principal for general c is

$$P''(a^P) = -\int_m^M c_{aa}(x,a^P) \cdot f(x)\, dx$$
$$+ (s^1 + s^2) \cdot f(a^P) < 0 \;. \qquad (5.17)$$

In our case this is

$$-\int_m^M 2 \cdot f(x)\, dx + (s^1 + s^2) \cdot f(a^P) < 0 \;, \qquad (5.18)$$

i.e.,

$$(s^1 + s^2) \cdot f(a^P) < 2 \;. \qquad (5.19)$$

For each bounded density functions there exist small values s^1 and s^2 such that condition (5.18) is satisfied. Since the principal does not know the function f he is kind of trapped: higher values for the s^i increase his expected income, but if they are "too high" he might end up with a minimum.

This result, which admittedly is absurd at first sight, has a serious economic foundation: If the fine for differences between the agent's reported value and true output is high enough then the principal on average benefits from the agent's misprediction.

Condition (5.17) also tells us that, loosely speaking, more convexity of adjustment costs (an increase in c's second-order derivative with respect to a) allows higher values for the s^i without violating the second-order condition.

If instead of (5.1) the payment function

$$\tilde{y}(x,a) = w \cdot x - (a - x)^2 \qquad (s > 0) \qquad (5.20)$$

is chosen by the principal, the agent has to maximize

$$\tilde{A}(a) = \int_m^M [w \cdot x - (a - x)^2] \cdot f(x)\, dx$$
$$= w \cdot E(x) - a^2 + 2 \cdot a \cdot E(x) - \int_m^M x^2 \cdot f(x)\, dx . \qquad (5.21)$$

A necessary condition for an interior solution of the agent's problem is

$$\tilde{A}'(a) = - 2 \cdot a + 2 \cdot E(x) = 0 . \qquad (5.22)$$

It is hence optimal to report the expected output. Obviously the solution derived from condition (5.22) indeed is a maximum. In this case the agent bears all the adjustment costs.

For the payment function (5.1) (choosing $s = s^1 = s^2$) and a symmetrical density function f, the agent will report the expected output E(x), and the principal's expected income is

$$P(a) = (p - w) \cdot E(x) - \int_m^M [x - E(x)]^2 \cdot f(x)\, dx$$
$$+ 2 \cdot s \cdot \int_m^{E(x)} [E(x) - x] \cdot f(x)\, dx . \qquad (5.23)$$

For the payment function $\tilde{y}$, however,

$$P(a) = (p - w) \cdot E(x) \qquad (5.24)$$

is obtained. Consequently it depends on the fine s (and on the density function f unknown to the principal) which of these payment functions he should prefer.

6. Research Agenda

As was pointed out earlier, a lot of important and interesting questions remain that altogether form a considerable research agenda. First, there is the question whether or not, respectively in which cases, optimal payment functions exist. And it is another problem to find ways to determine all such functions if they are not unique.

Third, it would be nice to know whether any or all such solutions are kind of second-best, i.e., given the principal's optimal payment function y^* the agent reports an optimal value a^* that does not maximize the principal's expected utility conditional on y^*.

Fourth, for practical applications a lot of further requirements seem sensible which pose additional difficulties: Guaranteeing a minimal income for the agent is necessary for acceptance of a payment function. Also it may be desired to only partially place the risk upon the agent. And an optimal payment function needs to be "sufficiently simple" to be understood by all agents.

Finally, dynamical aspects, too, would enrich the model. For example, shifting output or sales etc. into later periods should not increase income. This could be achieved by having prices fall from one period to another by something like a discount rate. Then, however, it has to be credible that the payment scheme (with different, initially known values for the parameters) will be applied for several periods.

Eventually, discussing the model and its refinements may even lead to inquiries into the meaning of honesty. And it is quite likely one will find that this is not just reporting truly some expected value.

The reader is obviously left with many puzzling questions. Bringing these questions to his attention that is what the paper is all about.

References

Bamberg, G. and F. Baur [1980], "Statistik," Oldenbourg, München.

Beckmann, M.J. [1987], "Managers as principals and agents," this volume.

Brechling, F. [1975], "Investment and employment decisions," Manchester University Press.

Gerhardt, G.R. [1967], "Struktur und Prozeßqualität materieller Anreizsysteme," Wissenschaftliche Zeitschrift der Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Reihe 16, 915-924.

Rees, R. [1985a], "The theory of principal and agent - Part I," Bulletin of Economic Research 37, 3-26.

[1985b], "The theory of principal and agent - Part II," Bulletin of Economic Research 37, 75-95.

Weitzman, M. [1976], "The new Soviet incentive model," Bell Journal of Economics 7, 251-257.

Section 3 **Capital Markets and Moral Hazard**

Moral Hazard and Equilibrium Credit Rationing: An Overview of the Issues

Helmut Bester and Martin Hellwig

Summary: One of the more intriguing puzzles in microeconomics is presented by the phenomenon of credit rationing. If funds are so scarce as to require rationing, why do lenders not raise the interest that they demand? We survey recent developments that seek to explain this phenomenon by appealing to incentive problems in the relation between the borrower and the lender. A simple example, due to Stiglitz and Weiss, shows that under certain circumstances, lenders will not use their bargaining power to raise interest rates because the adverse incentive effects of such a move outweigh any direct effect on the lender's payoffs. To examine the robustness of this argument, we discuss how the analysis is affected by the use of collateral, variations in loan size and investment, or alternative forms of the finance contract. Finally, we analyse the relation between the credit-rationing problem and the general theory of optimal incentive schemes under imperfect information.

Table of Contents:

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

1. Introduction

A would-be borrower is said to be rationed if he cannot obtain the loan that he wants even though he is willing to pay the interest that the lenders are asking, perhaps even a higher interest. In practice such credit rationing seems to be commonplace: Some borrowers are constrained by fixed lines of credit which they must not exceed under any circumstances; others are refused loans altogether. As far as one can tell, these rationing phenomena are more than the temporary consequences of short-term disequilibrium adjustment problems. Indeed they seem to inhere in the very nature of the loan market.

For the ordinary microeconomist, such rationing phenomena present a puzzle. The equilibrium of a market is commonly identified with the balance of demand and supply. According to the law of demand and supply, prices in the market should adjust until any excess of demand over supply or of supply over demand has been eliminated, at which point there is no more room for rationing. By this logic, any credit rationing should be accompanied by increases of interest rates that reduce the demand for loans and raise the supply of loans until the need for credit rationing has disappeared.

The law of demand and supply is usually justified by the more general principle that economic agents act in their own perceived self-interests. An excess supply or excess demand would enable the agents on the short side of the market to move prices in a direction which makes them better off. Thus a seller should be expected to exploit excess demand by charging higher prices.

The argument against rationing as an equilibrium phenomenon is to some extent independent of the market structure. While the law of demand and supply has been proposed for competitive markets, the underlying behavioural principle may be applied to monopolistic markets as well. A monopolist too will prefer to raise his prices rather than ration demand at given, low prices.

Given the general principle that rationing is at most a transitory disequilibrium phenomenon, economists have found it difficult to come to terms with the phenomenon of credit rationing. In many cases of course, credit rationing can be explained by government interference with the market: Usury laws, interest rate and bank regulation, and certain types of central bank intervention. However, there has always been a suspicion that this is not the whole story. Beginning with Hodgman (1960), a series of papers in the early sixties discussed the possibility that credit may be rationed because a lender does not want to grant a loan that exceeds the borrower's ability to repay. This observation was soon found to be besides the point because a borrower typically does not want to have a loan that he knows he cannot repay (for an excellent discussion of these issues, see Clemenz (1986), Chapter 1). The deeper problem of credit rationing relative to what the borrower wants was not addressed by this literature; indeed this problem remained unsolved for a long time.

In recent years, economists have tried to relate the phenomenon of credit rationing to problems of imperfect information. Such problems arise when the lender tries to evaluate the borrower's promise of repayment at some later date. The quality of this promise depends on the behaviour and the characteristics of the borrower. In both respects, the borrower typically has private information. Thus an entrepreneur may have better information than his bank about the objective prospects of his enterprise. At the same time, he is in a better position to control the risks that he takes or the amount of effort that he puts into his firm. All these factors affect the value of the lender's claim, and yet he is unable to control them directly.

In this situation, the lender must take account of the effects of the credit contract on the mix of loan applicants or on their behaviour. An increase in interest rates might lead borrowers with fairly safe projects to drop out of the market, or it might induce them to replace their safe projects by riskier ones. Such considerations may cause a lender to refrain from raising interest rates even though he has the bargaining power to do so.

The incomplete information approach to interest rigidity and credit rationing was first developed by Jaffee and Russell (1976), Keeton (1979), and Stiglitz and Weiss (1981). In particular, Stiglitz and Weiss (1981) show that credit rationing can be an equilibrium phenomenon if either the lender is imperfectly informed about the borrower's characteristics or the lender is unable to directly control the borrower's behaviour. In the following we discuss the latter phenomenon where credit rationing is a consequence of moral hazard in the borrower-lender relationship. In view of the extensive surveys by Baltensperger and Devinney (1985), as well as Clemenz (1986), we do not aim for completeness in our treatment of the literature. Instead, we shall discuss the original Stiglitz-Weiss example and look at several modifications in order to see which structural elements of the example are crucial. At the same time, we propose to relate the theory of credit rationing under moral hazard to the general theory of incentive problems as treated e.g. by Grossman and Hart (1983).

2. Moral Hazard and Equilibrium Credit Rationing: The Leading Example

2.1 Loan Contracts and Risk Taking

Consider an entrepreneur who can choose between two investment projects, indexed $i=a,b$. Both projects require the same fixed investment I . The returns to both projects are risky; for $i=a,b$, project i earns the return

$$(1)\qquad \tilde{X}_i = \begin{cases} X_i & \text{with probability } p_i \\ 0 & \text{with probability } 1-p_i \end{cases},$$

where

$$(2)\qquad p_a X_a > p_b X_b > I,\ 1 > p_a > p_b > 0,\ X_b > X_a \ .$$

For simplicity, both projects have only two possible outcomes, success and failure. Project a is more likely to succeed, but project b has the higher return in the case of success. In the case of failure, neither project yields anything. Project a has the higher expected return, but even project b's expected return exceeds the cost I .

The entrepreneur has no initial wealth. He uses debt finance to undertake the investment. A debt contract is characterized by a gross interest payment R which the entrepreneur must pay the lender in the case of success. If the project fails, the entrepreneur goes bankrupt, and the lender receives nothing. Given the interest payment R , the entrepreneur's expected payoff from undertaking project i is given as

$$(3)\qquad U_i(R) = p_i(X_i - R) \ .$$

The entrepreneur is taken to be risk neutral so that he applies for a loan as long as his expected payoff is nonnegative.

Lenders, too, are taken to be risk neutral. Given the contractual interest payment R , a lender's expected payoff from financing the entrepreneur's investment in project i is given as

$$(4)\qquad \pi_i(R) = p_i R - I \ .$$

Given R, the lender prefers the entrepreneur to undertake the project with the higher success probability. Under perfect information, the loan contract would therefore prescribe not only the interest payment R, but also the choice of the project i that is to be undertaken.
However we assume that the relation between the entrepreneur and any lender is subject to moral hazard because the lender cannot observe the entrepreneur's choice of project. Therefore the loan contract cannot effectively prescribe the project that is to be undertaken. The loan contract can only specify the interest R which the entrepreneur pays if his project - whichever one he chooses - happens to succeed.

Given the interest obligation R, the entrepreneur selects the project which maximizes his expected payoff. As Stiglitz and Weiss (1981) have observed, this decision depends on R. From (3), the entrepreneur is willing to choose project a if and only if

(5) $$p_a(X_a-R) \geqq p_b(X_b-R) \quad .$$

If we write $i(R)$ for the entrepreneur's project choice under a contract with interest payment R, we see that there is a critical level

(6) $$\hat{R} := \frac{p_a X_a - p_b X_b}{p_a - p_b}$$

such that

(7) $$i(R) = \begin{cases} a & \text{if } R < \hat{R}, \\ b & \text{if } R > \hat{R}. \end{cases}$$

As R rises above $\hat{R}$, the entrepreneur switches from project a to project b, which has the higher probability of failure. Quite generally, high interest obligations lower the entrepreneur's payoff in the case of success and reduce his incentives to avoid bankruptcy.
For $R = \hat{R}$, the entrepreneur is indifferent between the two projects. For simplicity, we assume that in this case he chooses project a, i.e. we set $i(\hat{R}) = a$.
Lenders must take account of the effects of R on the entrepreneur's behaviour. Given a lender's inability to monitor the entrepreneur's project choice, his expected payoff from a contract with interest payment R is

(8) $$\pi^*(R) = \pi_{i(R)}(R) = \begin{cases} p_a R - I & \text{if } 0 \leqq R \leqq \hat{R}, \\ p_b R - I & \text{if } \hat{R} < R \leqq X_b. \end{cases}$$

The form of $\pi^*(.)$ is illustrated in Figure 1. Since $p_a > p_b$, $\pi^*(.)$ is <u>not</u> monotonically increasing in R. At $R = \hat{R}$, any small increase in the interest payment leads to a discontinuous drop in $\pi^*(R)$ as the entrepreneur switches to the project with the higher bankruptcy probability. This nonmonotonicity of the lender's expected payoff function is the basis for the theory of credit rationing proposed by Stiglitz and Weiss (1981).

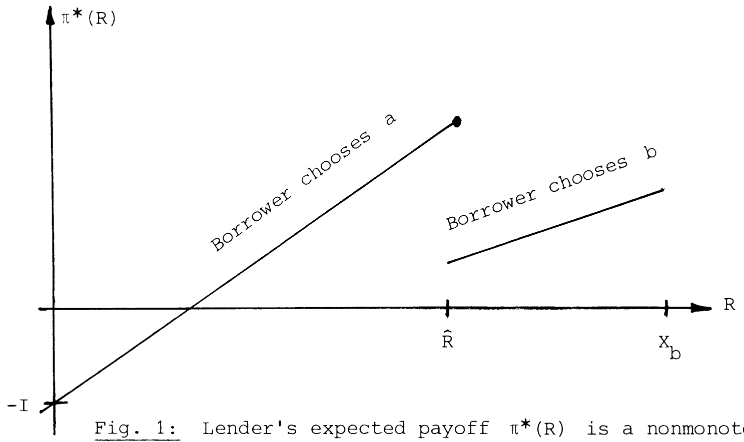


Fig. 1: Lender's expected payoff $\pi^*(R)$ is a nonmonotonic function of the contractual interest payment R.

2.2 Equilibrium Credit Rationing

According to Stiglitz and Weiss (1981), <u>credit rationing occurs when some loan applicants receive loans and others do not, although the latter would accept even higher interest payments</u>. We now show that even a credit market equilibrium may involve rationing when there is moral hazard.

We first consider the case of a monopolistic loan market. Suppose that there is a single risk neutral lender who owns an amount L of loanable funds. Furthermore suppose that there are N identical entrepreneurs of the type described above, and let $I \leqq L < N\,I$. Then funds are scarce, and the lender is unable to finance all entrepreneurs.

In this situation, the lender has all the bargaining power. He can set the terms of the contract to maximize his return. In particular, he can impose an interest obligation R^* at which the value of his expected payoff π^* is maximal. By inspection of (8), there are two possibilities for this choice. If

(9a) $\quad p_a \hat{R} < p_b X_b$,

then $\pi^*(\hat{R}) < \pi^*(X_b)$, and the lender's payoff is maximized at $R^* = X_b$. Alternatively, if

(9b) $\quad p_a \hat{R} > p_b X_b$,

then $\pi^*(\hat{R}) > \pi^*(X_b)$, and the lender's payoff is maximized at $R^* = \hat{R}$. (If $p_a \hat{R} = p_b X_b$, the lender's payoff is maximized both at $R^* = \hat{R}$ and at $R^* = X_b$; in what follows, we neglect this case.)

Under the parameter constellation (9b), we must have equilibrium credit rationing. The lender announces the contractual interest payment $R^* = \hat{R}$ which maximizes his payoff expectation. Given the interest obligation $\hat{R}$, an entrepreneur who gets a loan can expect the payoff $U^*(\hat{R}) = U_{i(\hat{R})}(\hat{R}) = p_a(X_a - \hat{R})$. By inspection of (6) and (2), we have

(10) $\quad U^*(\hat{R}) > 0$,

i.e., any entrepreneur has a strict preference for undertaking the investment. Therefore all entrepreneurs apply for loans, and the lender must somehow select L/I applicants to distribute his funds. The remaining $N - L/I$ applicants are rejected and envy their colleagues who undertake their investments and earn positive profits. Indeed any applicant who is rejected would gladly offer to pay more than $\hat{R}$ in order to get a loan. However the lender will refuse such an offer because it would effectively make him worse off.

It may be useful to compare the credit market equilibrium under the parameter constellation (9b) with the equilibrium under the parameter constellation (9a). Under the parameter constellation (9a), the lender announces the required interest payment $R^* = X_b$. Given this announcement, an entrepreneur who gets a loan can expect the payoff $U^*(X_b) = U_{i(X_b)}(X_b) = p_b(X_b - X_b) = 0$, i.e., any entrepreneur is indifferent about whether he undertakes his investment or not. As before, the lender provides loans to L/I entrepreneurs, leaving $N - L/I$ entrepreneurs without funds. However, in this case an entrepreneur who fails to get a loan does not envy his colleagues; moreover he is neither willing nor able to pay more than the announced interest payment $R^* = X_b$.

Under the parameter constellation (9a), there is thus no credit rationing in equilibrium. To be sure, in equilibrium, some would-be borrowers receive loans and others do not. However the latter are just as happy as the former because they all receive the same payoff. In contrast, under the parameter constellation (9b), those entrepreneurs who receive loans are strictly better off than the others.

In order to see more clearly the connection between rationing and moral hazard, it is also helpful to consider the equilibrium which emerges under perfect information. In this case, the only restriction that the lender has to observe is that the terms of the contract be acceptable to the borrowers. He can monitor the behaviour of firms and determine the choice of project. Given the scarcity of funds, he appropriates the entire surplus. Given that he appropriates the entire surplus, he asks that project a be undertaken because by (2), it yields the higher expected return. The interest payment to the lender is fixed at $R = X_a$.

Under perfect information again, the scarcity of funds does not entail rationing. As before, the lender finances L/I entrepreneurs. Each of these entrepreneurs undertakes project a and receives the payoff $U_a(X_a) = 0$, the same as what he would get without a loan.

Under imperfect information, this outcome is no longer feasible. If $R = X_a$, any borrower will switch to project b by which he obtains $U_b(X_a) = p_b(X_b-X_a) > 0$. Thus the lender can no longer do both, extract the entire surplus <u>and</u> implement project a at the same time. Given that he must choose between these alternatives, under the parameter constellation (9b), he prefers to implement project a even though this requires him to leave some of the surplus to the borrower. More generally, under imperfect information, one may find it more important to induce cooperative behaviour from one's partner than to appropriate the entire surplus from the partnership.

The phenomenon of equilibrium credit rationing is not limited to the case of a monopolistic loan market. Rationing may also occur when there are many lenders and the supply of funds is variable. To demonstrate this, consider an aggregate (competitive) supply function $L(.)$ for loanable funds. $L(.)$ may be taken to be an increasing function of the lenders' rate of return π/I so that for π sufficiently large, it may well be the case that $L(\pi/I) > NI$. In this case there are at least potentially enough funds for all firms to undertake the investment project. Nevertheless, under the parameter constellation (9b), this market has a credit rationing equilibrium if

(11) $L(\pi^*(\hat{R})/I) < NI$.

The equilibrium loan contract specifies the interest payment R so that lenders receive the rate of return $\pi^*(\hat{R})/I$. At this rate of return, the supply of funds is too small to satisfy total demand so that some entrepreneurs must go without loans. As before, the entrepreneurs who do not get loans envy those who do, and we have credit rationing.

To see that this outcome indeed constitutes an equilibrium, we note that none of the lenders has any incentive to deviate from it under any circumstances. The rate of return $\pi^*(R)/I$ that lenders receive is already the highest rate that is at all achievable in the market. Moreover at this rate of return, lenders lend out all the funds that they want to lend out. Those entrepreneurs who are denied credit will therefore find it impossible to change the situation. As in the monopoly case, we have a credit rationing equilibrium because (i) under the parameter constellation (9b), lenders achieve the highest return at the interest payment $\hat{R}$ at which borrowers have strictly positive payoff expectations, and (ii) at the rate of return $\pi^*(\hat{R})/I$, the supply of funds falls short of the demand.

How robust is the preceding analysis to changes in the basic model? In the following, we consider several modifications and extensions of the simple example that we have used so far. Our purpose is to determine more precisely which of the specific features of the example are responsible for the occurrence of equilibrium credit rationing.

3. Extensions and Modifications of the Analysis

3.1 Collateral as an Incentive Device

In addition to the assumptions of Section 2, we now suppose that each entrepreneur is endowed with some amount W of collaterizable wealth. This wealth cannot be used to finance investment directly, say because it consists of illiquid assets, or it represents the entrepreneur's future outside income. However, this wealth may be used as collateral for a loan. A loan contract then specifies not only a required interest payment R , but also a collateral $C \leq W$. The borrower loses C when he goes bankrupt. Accordingly, (3) has to be modified, and the entrepreneur's expected payoff from undertaking project i under a contract (R,C) becomes

$$(12) \qquad V_i(R,C) = p_i(X_i-R) - (1-p_i)C$$
$$= U_i(R-C) - C \quad .$$

The lender's valuation of C is not necessarily the same as the borrower's. Taking possession of collateral and liquidating it typically involves transactions costs. For simplicity, these will be represented by a factor $1-\beta$, with $0 \leqq \beta \leqq 1$, so that the lender's evaluation of C equals βC. The lender's expected payoff from financing the entrepreneur's investment in project i through a contract (R,C) is therefore given as

$$(13) \qquad \varphi_i(R,C) = p_i R + (1-p_i)\beta C - I$$
$$= \pi_i(R-\beta C) + \beta C \quad .$$

In the present context, collateral is not used as a means to enforce repayment. All along, we have assumed that contracts are enforceable, and that the borrower never defaults if his realized return permits repayment. In practice this willingness to repay the lender in the event of success may be motivated by the fact that the firm has been pledged as security for the loan. However, we are not concerned with such collateral inside the firm which is worth nothing when the firm fails. The collateral C that we consider here is an asset outside the firm which only comes into play when the firm fails so that its assets are worth nothing.

Under perfect information, such outside collateral should not play any significant role. From (12) and (13), it follows that any contract (R,C) with $C > 0$ is (weakly) dominated by another contract with $C = 0$. Indeed if $\beta < 1$, both the lender and the borrower can gain by reducing C to zero and increasing R appropriately because this operation yields a surplus of $(1-p_i)(1-\beta)C$. Thus the costs of collateralization will preclude its use under perfect information.

However, as shown by Bester (1985, 1987), under imperfect information, collateral may play a significant role. In the present context, the lender may use the collateral requirement to influence the entrepreneur's project choice. Given a loan contract with the terms (R,C), the entrepreneur is willing to choose project a if and only if

$$(14) \qquad p_a(X_a-R) - (1-p_a)C \geqq p_b(X_b-R) - (1-p_b)C \quad ,$$

or

$$(15) \qquad R \leqq \hat{R} + C \quad ,$$

where $\hat{R}$ is defined as in (6). If we compare (15) with the previous incentive constraint $R \leqq \hat{R}$, we see that the use of collateral gives the lender more scope for inducing the entrepreneur to choose project a. In contrast to interest payments, collateral requirements have positive incentive effects. They effectively punish the borrower when his project fails, thus creating a motive to lower the probability of bankruptcy by choosing project a . For $\beta C \leqq R$, this incentive effect is favourable for the lender because it increases the probability of repayment.

The positive incentive effect of collateral requirements may induce the lender to impose such a requirement even if the transactions costs are high so that β is close to - or even equal to - zero. The point is that an increase in C gives the lender more room for increasing R without any adverse incentive effects. In particular, condition (15) shows that a simultaneous and equal increase in R and C will never have an incentive effect at all. From (13) it follows that such an equal increase in R and C will unambiguously raise the lender's payoff - even if $\beta = 0$ so that the collateral does not actually enter the lender's receipts directly.

Because of its incentive effects, the use of collateral requirements substantially affects the scope for equilibrium credit rationing. As before, we consider the case of a monopolistic lender whose funds are insufficient to satisfy all the borrowers' needs. The lender can impose an interest payment R^* and a collateral requirement $C^* \leqq W$ subject only to the constraint that the borrower's expected payoff should not be negative. Again the lender must choose whether he wants to implement project a or project b . If he decides to implement project b , he can appropriate the entire surplus of the enterprise, e.g. by setting $R^* = X_b$, $C^* = 0$, for an expected payoff $p_b X_b - I$. If he wants to implement project a , it is most profitable for him to set $R^* = \hat{R} + C^*$, the maximum compatible with condition (15), and to set $C^* = \min[W, p_a(X_a - \hat{R})]$, the maximum compatible with both the constraints $C^* \leqq W$ and $V_a(\hat{R}+C^*, C^*) \geqq 0$.

We must now distinguish three possible parameter constellations. If

(16a) $$p_a \hat{R} + [p_a + (1-p_a)\beta] \min [W, p_a(X_a - \hat{R})] < p_b X_b \quad ,$$

the lender prefers to implement project b and to appropriate the entire surplus of the project by setting $R^* = X_b$, $C^* = 0$. As before, this case does not involve credit rationing because the loan applicants are indifferent about whether they receive loans or not.

Alternatively, if

$$\text{(16b)} \quad \begin{cases} W < p_a(X_a-\hat{R}) & \text{and} \\ p_a \hat{R} + [p_a + (1-p_a)\beta]W > p_b X_b , \end{cases}$$

the lender finds it most profitable to set $R^* = \hat{R} + W$ and $C^* = W$, thus implementing project a . In this case, the borrower's payoff is $V_a(\hat{R} + W,W) = p_a(X_a-\hat{R}) - W$, which is strictly positive. The insufficiency of the lender's funds leads to equilibrium credit rationing because again the loan applicants who are rejected envy those who are accepted and would gladly offer to pay more than the interest $\hat{R} + W$ that the lender is asking.

Finally, if

$$\text{(16c)} \quad \begin{cases} W \geqq p_a(X_a - \hat{R}) & \text{and} \\ p_a \hat{R} + [p_a + (1-p_a)\beta]p_a(X_a-\hat{R}) > p_b X_b , \end{cases}$$

the lender again wants to implement project a , this time however by setting $R^* = p_a X_a + (1-p_a)\hat{R}$ and $C^* = p_a(X_a-\hat{R})$. The borrower's expected payoff then is zero, i.e. the individual rationality constraint $V_a(R^*,C^*) \geqq 0$ is binding. Even though some loan applicants are rejected, the equilibrium does not involve rationing because those loan applicants who are rejected do not care and are unwilling to offer more than the lender is asking.

To assess the impact of collateral requirements on the possibility of equilibrium credit rationing, we compare condition (16b) with condition (9b), our previous condition for equilibrium credit rationing. Obviously the two conditions coincide if $W = 0$. For $W > 0$, we must distinguish two possibilities: If W is very high, equilibrium credit rationing is impossible because the use of collateral enables the lender to appropriate the entire surplus from project a in an incentive-compatible way. However, if $p_b X_b > p_a \hat{R}$ and if W lies in some intermediate range, the use of collateral may actually <u>cause</u> equilibrium credit rationing as it becomes more profitable for the lender to implement project a and to replace the contract $(X_b,0)$ by $(\hat{R}+W,W)$.

The different possibilities are illustrated graphically in Figure 2. In this figure, the line AA represents the equation $W = p_a(X_a-\hat{R})$, or $V_a(\hat{R}+W,W) = 0$. For parameter constellations above this line, there never is any credit rationing because the lender can always use collateral to push the borrower to the point where he is indifferent about borrowing at all. Contour BBB represents the equation

$p_b X_b = p_a \hat{R} + [p_a + (1 - p_a)\beta] \min[W, p_a(X_a - \hat{R})]$. To the right of this contour, the lender prefers to implement project b , to the left, project a .

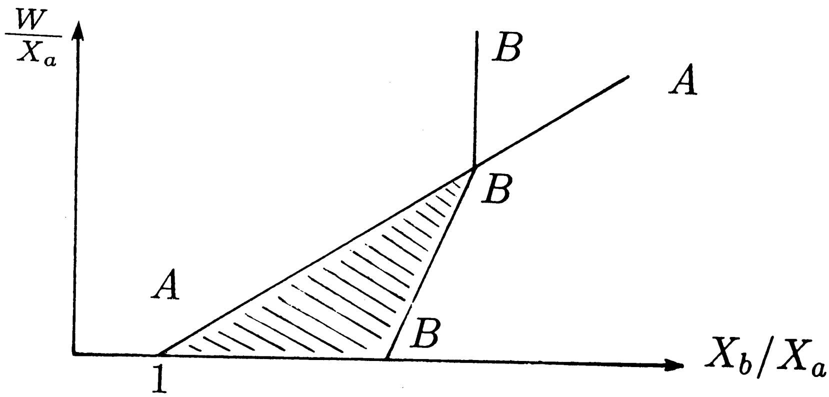


Fig. 2: Credit rationing occurs only in the hatched triangle in which the lender wants to implement project a and the borrower's collateralizable wealth W is insufficient for the lender to appropriate all the surplus.

It is of some interest to note that e.g. for $p_b X_b > p_a \hat{R}$, the equilibrium choice of project depends on the borrower's wealth. This contrasts with the well-known Modigliani-Miller theorem in corporate finance according to which the firm's production and financial decisions are independent of each other. Under imperfect information, this theorem fails because finance contracts have incentive effects which are relevant for production decisions.

3.2 Credit Rationing and the Level of Investment

We now abandon the assumption that the level of investment in each project is fixed. If the level of investment in each project is fixed and if funds are scarce, then necessarily some loan applications must be rejected. If investment levels are variable, it may also be possible to undertake <u>all</u> projects on a reduced scale. In the following, we consider the implications of this possibility for the theory of credit rationing.

For this purpose, we modify the specification of returns in Section 2. Let $f(.)$ be a standard "S-shaped" production function which exhibits (possibly) first increasing and then decreasing returns to scale, i.e. we have

(17a) $f(0) = 0$, $f'(I) > 0$ for all I

and moreover there exists a minimum efficient scale $I^* \geqq 0$ such that

(17b) $f(I^*) > I^* \frac{f(I)}{I}$

for any $I \in (0, I^*)$, and

(17c) $f''(I) < 0$

for any $I \geqq I^*$.

We replace the return specification (1), (2) by the new specification

(18) $$\tilde{x}_i = \begin{cases} X_i\, f(I) & \text{with probability } p_i \\ 0 & \text{with probability } 1-p_i \end{cases} ,$$

where p_a, X_a, p_b, X_b satisfy (2), and moreover

(19) $f(I^*) \geqq I^*$ and $f'(I^*) > 1$.

The project return is now related to the level of investment, and the market equilibrium must determine the amount of funds which is allocated to each project.

In all other respects, the model of Section 2 is unchanged. In particular, we return to the case where the entrepreneur has no collateralizable wealth.

Within the extended framework, the investment level becomes part of the credit contract. When the entrepreneur has obtained a contract (R,I) , his payoff from undertaking project i is given as

(20) $U_i(R,I) = p_i[X_i f(I) - R]$.

The lender's payoff from financing the entrepreneur's investment in project i through a contract (R,I) is given as

(21) $\pi_i(R,I) = p_i R - I$,
as in formula (4) .

Under imperfect information, both the interest payment R and the investment level I have incentive effects. The borrower's choice of project is determined by considerations which are analogous to (5). He is willing to choose project a if and only if the terms of the loan contract satisfy the incentive constraint

(22) $R \leqq \hat{R} f(I)$,

where again $\hat{R}$ is defined as in (6). If the lender wishes to implement project a, condition (22) constrains the set of contracts that are available to him.

What are the implications of (22) for the possibility of equilibrium credit rationing? We consider this question first for the case $I^* = 0$ of a production function with everywhere decreasing marginal returns. In this case, credit rationing in the sense of Stiglitz and Weiss is no longer compatible with equilibrium. Equilibrium credit rationing requires that the incentive constraint (22) be binding since otherwise the lender could raise R without any incentive effects. At the contract (R,I) satisfying $R = \hat{R} f(I)$, the lender earns the rate of return $[p_a\hat{R} f(I) - I]/I$. By (17c) and the assumption that $I^* = 0$, a reduction in I increases $[p_a \hat{R} f(I) - I]/I$. This means that the lender could gain by slightly reducing R and I in such a way that (22) continues to hold with equality. Due to the reduction in I , he needs fewer funds on the given population of borrowers, i.e. he has a surplus to use for whatever borrower was previously rationed. As long as there are any entrepreneurs who do not receive loans at all, the lender can improve his payoff by this policy of reducing both the individual loan and the interest payment in order to spread the available funds more evenly among the population of loan applicants. Thus if $I^* = 0$, rationing in the sense of Stiglitz and Weiss cannot persist in equilibrium.

More generally, if there are N identical entrepreneurs and an amount L of loanable funds in the market, there cannot be any equilibrium credit rationing if $L > N I^*$. If $L > N I^*$, then under (17c) and (19), the highest investment that any firm undertakes in equilibrium must exceed I^* and lie in the region of diminishing marginal returns. In this region a smaller loan size and a smaller investment level have positive incentive effects. Specifically, for $I > I^*$ a reduction in R and I by the same percentage rate increases the entrepreneur's payoff per unit of investment. Consequently this measure motivates the choice of safer projects. By appropriately setting R and I the lender can avoid incentive problems and increase his profit as long as there is excess demand.

In contrast, equilibrium credit rationing is possible if $I^* > 0$ and $L < N I^*$ so that it is impossible to provide all loan applicants with the minimum efficient scale I^* . In this case, the lender will choose to reject some loan applicants in order to provide the others with loans of size I^* rather than to provide all loan applicants, say, with equal loans of size L/N . At any investment level I , the lender's maximal rate of return is $[p_a \hat{R} f(I) - I]/I$ if he wants to implement project a (through $R = \hat{R} f(I)$) and $[p_b X_b f(I) - I]/I$ if he wants to implement project b (through $R = X_b f(I)$). In either case his rate of return is maximal at $I = I^*$, the minimum efficient investment scale. If $L < N I^*$, it is therefore more profitable to make L/I^* loans, each of size I^*, rather than N loans of size L/N .

As in Section 2, the possibility of equilibrium credit rationing now turns on inequalities (9a) and (9b). If $p_a \hat{R} < p_b X_b$, the lender prefers to implement project b , i.e. he provides L/I^* entrepreneurs with the loan contract $(X_b f(I^*), I^*)$ at which they are just indifferent about their projects altogether. As in Section 2, this constellation does not involve rationing. However, if $p_a \hat{R} > p_b X_b$, the lender prefers to implement project a by providing the contract $(\hat{R} f(I^*), I^*)$. In this case, the $N-L/I^*$ entrepreneurs who do not get a loan envy the L/I^* ones who do, and would be willing to offer a strictly higher interest rate. We conclude that with a variable investment level equilibrium credit rationing is possible if and only if the minimum efficient scale of investment is strictly positive and moreover the available funds are insufficient to finance the minimum efficient scale for all loan applicants.

When the level of investment in each project is variable, credit rationing can also take the form of <u>loan size rationing</u> in the sense that <u>some loan applicants receive smaller loans than others even though they</u>

<u>would be willing to accept even higher interest obligations than the others</u>. To show this, we consider a production function f with multiple locally efficient scales of investment. Specifically, we replace (17b,c) by the condition that for some $I^* > 0$ and any $k=0,1,2,...,$

$$(23a)\qquad \frac{f((k+1)I^*)}{(k+1)I^*} = \frac{f(I^*)}{I^*}$$

and moreover,

$$(23b)\qquad \frac{f(I)}{I} < \frac{f((k+1)I^*)}{(k+1)I^*}$$

for any $I \in (k\,I^*,(k+1)I^*)$. Conditions (23a) and (23b) imply that the lender's rate of return is maximal for loans of sizes $I^*, 2\,I^*, 3I^*, \ldots$. Moreover for loan sizes that are not integer multiples of I^* , his rate of return is strictly less.

Now suppose that the total amount of available funds L is such that $N\,k\,I^* < L < N(k+1)I^*$ for some fixed k . In this case the lender may choose to make $N \cdot \frac{(k+1)I^* - L/N}{I^*}$ loans of size $k\,I^*$ and $N\,\frac{L/N - k\,I^*}{I^*}$ loans of size $(k+1)I^*$. Under this policy every applicant receives a loan, but still the scarcity of funds and the locally increasing returns to scale induce the lender to provide different loans to different applicants.

As before, the possibility of rationing as an equilibrium phenomenon turns on inequality (9b). If $p_a \hat{R} > p_b X_b$, the lender wants to implement project a in all contracts. In view of the incentive compatibility constraint (22), he imposes the interest obligations $\hat{R}\,f(k\,I^*)$ and $\hat{R}\,f((k+1)I^*)$ on loans of sizes $k\,I^*$ and $(k+1)I^*$. Under these terms, an entrepreneur's payoff is

$$(24a)\qquad U_a(\hat{R}\,f(k\,I^*),I^*) = p_a(X_a-\hat{R})\;f(k\,I^*)$$

if he gets a loan $k\,I^*$ and

$$(24b)\qquad U_a(\hat{R}\,f((k+1)I^*),\,(k+1)I^*) = p_a(X_a-\hat{R})\;f((k+1)I^*)$$

if he gets a loan $(k+1)I^*$. Since $X_a > \hat{R}$ and $f((k+1)I^*) > f(k\,I^*)$, we have

$$(24c)\qquad U_a(\hat{R}\,f((k+1)I^*),\,(k+1)I^*) > U_a(\hat{R}\,f(k\,I^*),\,k\,I^*)\quad ,$$

i.e. any entrepreneur who receives the smaller loan $k\,I^*$ is strictly envious of those entrepreneurs who receive the larger loan $(k+1)I^*$. To get an increment I^* to his loan, he would be willing to pay strictly more than the difference in interest, $\hat{R}[f((k+1)I^*) - f(k\, I^*)]$, that the lender is asking. However, for reasons which are by now familiar, the lender is not willing to accept such an offer, and we have equilibria with credit rationing, this time in the form of loan size rationing.

Actually, the preceding example does not bring out the full significance of loan size as opposed to all-or-nothing rationing. In this example, the locally efficient investment scales $I^*, 2I^*, 3I^*, \ldots$ all yield the <u>same</u> rate of return so that the lender is exactly indifferent between a policy of loan size rationing and a policy of all-or-nothing rationing in the sense of Stiglitz and Weiss. Instead of making $N \frac{(k+1)I^* - L/N}{I^*}$ loans of size $k\, I^*$ and $N \frac{L/N - k\, I^*}{I^*}$ loans of size $(k+1)I^*$, the lender would be just as willing to make $L/h\, I^* < N$ loans of size $h\, I^*$, $h > k+1$, leaving $N - L/h\, I^*$ applicants without a loan altogether.

This indifference between loan size rationing and all-or-nothing rationing disappears if one replaces conditions (23a) and (23b) by an assumption of globally decreasing returns to scale such that the locally efficient investment scales $I^*, 2I^*, 3I^*, \ldots$ satisfy the inequalities:

$$(25) \qquad \frac{f((k+1)I^*)}{(k+1)I^*} < \frac{f(k\, I^*)}{k\, I^*} \qquad \text{for } k=1,2,\ldots \, .$$

If (25) holds and if $N k I^* < L < N(k+1)I^*$ for some integer $k \geq 1$, then loan size rationing is the <u>only</u> possible form of credit rationing. As long as there is any loan applicant who does not receive credit at all, the lender will increase his payoff if he reduces the largest loan that he makes by I^* in order to make an additional loan of size I^* . A detailed analysis of optimal lending policies under condition (25) and more general production functions will be the topic of a subsequent paper.

Our notion of loan size rationing has little to do with what is commonly called loan size rationing or "Type I Credit Rationing" in the literature (see, e.g., Jaffee and Russell (1976) or Keeton (1979)). Under the usual definition, credit rationing in the form of loan size rationing ("Type I Credit Rationing") occurs when <u>lenders quote a constant interest rate on loans and then proceed to supply a smaller loan size than the borrowers demand at this interest rate</u>. This notion arises mainly from the view that interest rates play the same role for loan markets as prices do for Walrasian goods markets so that any deviation from "interest-rate-taking" should be regarded as rationing. When there is uncertainty, we consider the interpretation of the interest rate as a Walrasian price to be incorrect. In the absence of complete contingent securities markets (which would involve a vector of contingent claims prices), the interest rate is no more than a parameter of a complex contract in an incomplete markets system (Gale and Hellwig (1985)). Moreover there is no intrinsic reason why this parameter should be taken to be independent of the size of the loan.

Nevertheless it may be instructive to consider "Type I Credit Rationing" in the present context and to compare it to loan size rationing as we have defined it. The interest rate r which is implicitly associated with a debt contract (R,I) is given as:

$$r := \frac{R - I}{I} \quad .$$

Thus under conditions (23a), (23b), a loan $k\,I^*$ with interest obligation $\hat{R}\,f(k\,I^*)$ involves the implicit interest rate

$$r_k = \frac{\hat{R}\,f(k\,I^*) - k\,I^*}{k\,I^*} = \frac{\hat{R}\,f(I^*) - I^*}{I^*} \quad ,$$

independent of k . This contract involves "Type I Credit Rationing" because, as we have seen, an increase in the loan by I^* together with an increase in the interest obligation by $\hat{R}\,f((k+1)I^*) - \hat{R}\,f(k\,I^*) = (1+r_k)I^*$ is always desirable for the debtor. In contrast, the contracts $(X_b\,f(k\,I^*), k\,I^*)$ that the lender imposes if he wants to implement project b do not involve "Type I Credit Rationing". Here the implicit interest rate is

$$\hat{r}_k = \frac{X_b\,f(k\,I^*) - k\,I^*}{k\,I^*} = \frac{X_b\,f(I^*) - I^*}{I^*}$$

independent of k . For any investment level I and interest obligation $(1+\hat{r}_k)I$, the borrower's payoff would be

$$p_b\left[X_b\,f(I) - (1+\hat{r}_k)I\right] = p_b\,X_b\,I\left[\frac{f(I)}{I} - \frac{f(I^*)}{I^*}\right]$$

which by (23a,b) is nonpositive and indeed negative whenever I is not an integer multiple of I^* . At the constant interest rate $\hat{r}_k$, there is no investment level that the borrower prefers to the proposed level $k\,I^*$. Thus it appears that under conditions (23a), (23b), the incidence of "Type I Rationing" parallels almost exactly the incidence of credit rationing (all-or-nothing or loan size rationing) in our sense of the term.

Unfortunately, the preceding conclusion breaks down if we abandon the specification (23a), (23b). To see this, return to the original specification (17a)-(17c) and suppose that $L > N\, I^*$ so that there is no rationing in the sense of Stiglitz and Weiss. Each borrower receives a loan $L/N > I^*$; in the case where the lender wants to implement project b , the interest obligation is $X_b\, f(L/N)$, and the implicit interest rate is

$$\hat{r} = \frac{X_b\, f(L/N) - L/N}{L/N} .$$

At the contract $(X_b\, f(L/N),\, L/N)$, the entrepreneur's payoff is zero; at any contract $((1+\hat{r})I, I)$ with $I \in [I^*, L/N)$, his payoff would be

$$p_b \left[X_b\, f(I) - (1+\hat{r})I \right] = p_b\, X_b\, I \left[\frac{f(I)}{I} - \frac{f(L/N)}{L/N} \right] > 0 ,$$

i.e. at the interest rate $\hat{r}$, the borrower would actually prefer a <u>smaller</u> loan. The contract that maximizes the lender's payoff imposes a <u>reverse</u> "Type I Rationing" on the borrower.

In the case where the lender wants to implement project a , the borrower's interest obligation $\hat{R}\, f(L/N)$ entails the implicit interest rate

$$r = \frac{\hat{R}\, f(L/N) - L/N}{L/N} .$$

At the contract $(\hat{R}\, f(L/N),\, L/N)$, the borrower's payoff is $p_a(X_a - \hat{R})\, f(L/N) = p_b(X_b - \hat{R})\, f(L/N)$ regardless of whether he undertakes project a or project b . At any other contract of the form $((1+r)I, I)$, his payoff would be equal to

$$(26) \qquad \max_{i \in \{a,b\}} p_i \left[X_i\, f(I) - (1+r)I \right] = \max_{i \in \{a,b\}} p_i \left[X_i\, f(I) - \frac{\hat{R}\, f(L/N)}{L/N}\, I \right] .$$

If we denote by i^* the project chosen under the contract $((1+r)I, I)$, then the payoff (26) is easily seen to equal

$$(26') \quad p_{i^*}(X_{i^*} - \hat{R})\, f(L/N) + p_{i^*}\, X_{i^*}\, I \left[\frac{f(I)}{I} - \frac{f(L/N)}{L/N} \right] + p_{i^*}(X_{i^*} - \hat{R}) \frac{f(L/N)}{L/N} (I - L/N)$$

The first term in (26') is just the payoff from the actual contract $(\hat{R}\, f(L/N), L/N)$. The second term accounts for the fact that at a constant interest rate r which corresponds to the average investment return at L/N , the borrower may wish to <u>reduce</u> his investment to take advantage of the higher average investment return at $I < L/N$. The third term accounts for the fact that the lender appropriates only a fraction of the borrower's return at $(\hat{R}\, f(L/N),\, L/N)$ so that at a constant interest rate, the borrower may wish to <u>expand</u> his investment level.

The presence or absence of "Type I Rationing" now depends on the relative size of the second and third terms in (26'). If average products are almost constant (the elasticity of production, $\eta = If'(I)/f(I)$, is close to one), and if X_{i*} is significantly larger than $\hat{R}$, then the third term will outweigh the second so that, at a constant interest rate, the entrepreneur would prefer to choose $I > L/N$. Alternatively, if the average product of investment declines very steeply and if X_b is close to X_a so that

$$\hat{R} = X_a - (X_b - X_a) \frac{p_b}{p_a - p_b}$$

is close to both X_a and X_b, then the second term in (26') outweighs the third so that the entrepreneur would prefer an investment level $I < L/N$. We thus see that for different parameter constellations, the same specification (17a)-(17c) with $L > N I^*$ may involve "Type I Rationing" or reverse "Type I Rationing" even though there is no credit rationing in our sense.

For the more general case described through (25), the same type of analysis shows that even our notion of loan size rationing is compatible with both, "Type I Rationing" and reverse "Type I Rationing". At a constant interest rate, the borrower typically wants a different loan than the lender provides. The point is that the lender is in a position to dictate the terms of the contract, and he is well aware of the fact that the borrower's returns depend nonlinearly on his investment level. Given the overall scarcity of funds, the borrower's preference for a different investment level is as irrelevant as his preference for a lower interest obligation.

3.3 Share Finance

Up to now, we have assumed that entrepreneurs must finance their investments by issuing debt. This assumption seems to be quite restrictive. Following Hart (1985), we therefore examine the possibility of rationing when alternative financial instruments are used.

It is easy to see that in the example of Section 2, rationing as an equilibrium phenomenon disappears if entrepreneurs finance their projects by issuing shares rather than debt. A share contract is simply described by some number $\alpha \in [0,1]$, which specifies the division of the project return between the financier and the entrepreneur. When project i is chosen, the entrepreneur's expected payoff from such a contract is given by

(27) $(1-\alpha)\ p_i\ X_i$,

whereas the financier's payoff is given by

(27') $\alpha\ p_i\ X_i$.

From (27) and (27'),one immediately sees that the entrepreneur and his financier unanimously agree on the choice of a project. Given condition (2), either party's payoff is maximized by selecting project a . Thus the share contract eliminates the conflict of interest which arises under debt finance. As long as we are only concerned about project choice, this type of contract creates no incentive problems that would prevent financiers from reacting to excess demand by raising α . Consequently, in equilibrium there can be no rationing in the market for share contracts.

In general, there are several objections against this simple conclusion. First, a share contract cannot be implemented unless the financier is able to observe the ex-post realization of the borrower's return. To see this, consider the following information setting. Assume that the financier can neither monitor the choice of project nor observe the realized return X_i . He only knows whether the project has failed or not. Under these circumstances, a share contract may no longer be incentive compatible because the entrepreneur could adopt the following strategy: He chooses project b , but if the project succeeds he claims that project a has been chosen and pays $\alpha\ X_a$. Since the financier cannot prove the contrary, he cannot enforce the contractual payment $\alpha\ X_b$ rather than $\alpha\ X_a$. The entrepreneur's expected payoff under this strategy is

(28) $p_b(X_b - \alpha\ X_a)$

as opposed to $p_a(1-\alpha)\ X_a$ if he had chosen project a . Thus he is better off cheating whenever

(29) $$\alpha > \frac{p_a\ X_a - p_b\ X_b}{X_a(p_a - p_b)} ,$$

or $\alpha > \hat{R}/X_a$, where $\hat{R}$ is again given by (6) . If the financier wants to implement project a when he cannot observe the value of X_i , he must respect the incentive-compatibility constraint

(30) $\alpha \leqq \hat{R}/X_a$,

which is exactly equivalent to the constraint $R \leqq \hat{R}$ in Section 2 .

More generally, Gale and Hellwig (1985) have shown that the standard debt contract is the optimal incentive-compatible form of finance when financiers cannot observe an entrepreneur's return realizations without costs. This general result provides some justification for our seemingly arbitrary initial restriction to debt finance.

Secondly, the use of share contracts may also be limited by other sources of moral hazard. Whereas share contracts dominate debt contracts when the entrepreneur's project choice is the only matter of concern, the reverse may be true when the entrepreneur's return depends on his effort in managing the firm (see, e.g., Jensen and Meckling (1976)). In this case, a share contract induces the entrepreneur to choose a sub-optimal level of effort because he receives only a fraction of the marginal return from additional effort. Jensen and Meckling (1976) have therefore suggested that in the presence of moral hazard involving both the entrepreneur's project choice and his effort level, an optimal finance contract would involve some mixture of debt and equity finance. In the following, we investigate what actually is an optimal finance contract in such a situation of "double moral hazard" and what are the implications of such double moral hazard for the possibility of rationing as an equilibrium phenomenon.

4. Rationing in the General Theory of Optimal Contracts under Moral Hazard

4.1 Optimal Finance Contracts and Rationing in a Model of Double Moral Hazard

In this last part of our survey, we relate the problem of credit rationing under moral hazard to the general theory of incentive problems. We start from the foregoing observation that under certain circumstances, the possibility of equilibrium credit rationing depends on which financial instruments are available. As we suggested above, this observation is not a general one. In some cases credit rationing as an equilibrium phenomenon may occur even though there are no restrictions at all on the form of the financial instruments that finance the entrepreneur's investment.

To substantiate this claim, we present an example with two sources of moral hazard. As before, there is a single financier with a limited supply of funds L. There are N entrepreneurs with investment projects that require a fixed outlay I. The return prospects of any one investment project now depend on the entrepreneur's effort ℓ as well as the project's risk class X. If the entrepreneur chooses the risk class $X \geqq 0$ and the effort level ℓ, his project earns the random return $\tilde{y}(X,\ell)$, where

$$(31) \qquad \tilde{y}(X,\ell) = \begin{cases} X\,g(\ell) & \text{with probability } p(X) \\ 0 & \text{with probability } 1 - p(X) \end{cases},$$

and $g(.)$ is an increasing, concave function, and $p(.)$ is a decreasing function.

As before, we assume that entrepreneurs have no wealth of their own. Moreover, $L < N\,I$ so that funds are scarce, and the financier has all the bargaining power. Thus he can set the terms of the finance contract as he likes. However, he is unable to control the entrepreneur's choice of X and ℓ. Therefore, he must take into account the incentive efforts of his contract choice on the entrepreneur's behaviour.

In contrast to the Gale-Hellwig analysis of debt contracts, we assume that the financier can costlessly observe the realization of the entrepreneur's return. Formally, a finance contract is therefore defined by a function $R(.)$ which specifies for each $y \geqq 0$ a payment $R(y) \leqq y$ that the entrepreneur must make to the financier if his return realization is equal to y. Given such a contract $R(.)$ and given the entrepreneur's choice of X and ℓ, the entrepreneur receives the expected payoff:

$$(32) \qquad U(X,\ell;R) = p(X)[X\,g(\ell) - R(X\,g(\ell))] - \ell ,$$

and the financier receives the expected payoff:

$$(33) \qquad \pi(X,\ell;R,I) = p(X)\,R(X\,g(\ell)) - I .$$

The financier's problem is to choose the contract $R(.)$ so as to maximize $\pi(X,\ell;R,I)$ subject to the constraints that for given $R(.)$, the entrepreneur chooses X and ℓ to maximize $U(X,\ell;R)$ and moreover that $U(X,\ell;R)$ must not be negative.

The formal analysis of the financier's problem is quite difficult because the incentive constraint on the choice of X and ℓ is not in general accessible to an analytic treatment. Formally, this incentive constraint amounts to a continuum of inequalities; namely, the chosen pair (X,ℓ) and <u>any</u> other pair (X',ℓ') that the entrepreneur might choose must satisfy the condition

$$(34) \qquad U(X,\ell;R) \geqq U(X',\ell';R)$$

(see Grossman and Hart (1983)). In the absence of any further assumptions, the continuum of inequalities (34) is analytically rather intractable.

4.2 Rationing and the Individual Rationality Constraint in a Parameterized Principal-Agent Problem

Because of this difficulty, we restrict our analysis to a special case in which the incentive constraints of the financier are easy to handle. Thus, we shall solve the financier's problem for the specification

$$(35) \qquad g(\ell) = \ell^{\beta}, \ 0 < \beta < 1 ,$$

$$(36) \qquad p(X) = e^{-X} .$$

Under the specification (35), (36), we can replace the incentive constraints (34) by the analytically more tractable first-order necessary conditions for the entrepreneur's maximization problem. Given a contract $R(.)$, the entrepreneur chooses X and ℓ to solve the problem

$$(37) \qquad \max_{X,\ell} \left\{ e^{-X}[X\ell^{\beta} - R(X\ell^{\beta})] - \ell \right\} .$$

In the absence of any information about the function $R(.)$, it is useful to decompose this problem into two stages. At the first stage, the entrepreneur chooses a value y for his return realization in the event of success. At the second stage, he chooses a combination of X and ℓ that will achieve the desired value y. With this decomposition, it becomes obvious that the financier's choice of a contract R can affect the entrepreneur's choice of y, the value of his return realization in the event of success. However, once y is given, the financier has no further influence on the entrepreneur's choice of the combination (X,ℓ) that will lead to the return realization y. In particular, the financier is exposed to the moral hazard that the entrepreneur may choose a combination of high risk X and low effort ℓ to obtain the given return realization y.

Formally, the second stage of our decomposition of the entrepreneur's problem requires that for any parametrically given y, he chooses X and ℓ to maximize (37) subject to the constraint $X\ell^{\beta} = y$. Any solution to this problem must satisfy the necessary first-order condition

(38) $\quad - e^{-X}[X\ell^{\beta} - R(X\ell^{\beta})] + \ell/\beta X = 0$

as well as the constraint $X\ell^{\beta} = y$. Thus, if the financier wants the entrepreneur to choose a pair (X^*,ℓ^*) , he must impose a contract $R(.)$ such that

(38') $\quad R(X^*\ell^{*\beta}) = X^*\ell^{*\beta} - \dfrac{\ell^* e^{X^*}}{\beta X^*}$.

However, whereas (38') is necessary for the implementation of X^*,ℓ^*, for several reasons it is <u>not</u> sufficient:

- Given $y^* = X^*\ell^{*\beta}$, (38') guarantees only that (X^*,ℓ^*) is a critical point for the problem of maximizing (37) subject to the constraint $X\ell^{\beta} = y^*$; to ensure that the entrepreneur chooses (X^*,ℓ^*) , this pair must correspond to a global maximum for the entrepreneur's second-stage problem.
- We have not yet ensured that the entrepreneur prefers (X^*,ℓ^*) over choices that yield return realizations $y \neq y^*$.
- We have not yet ensured that the entrepreneur's payoff at (X^*,ℓ^*) is nonnegative so that he accepts the contract at all.

We consider these issues in reverse order. Upon using (38') to substitute for $R(X^*\ell^{*\beta})$ in (37), we see that if (38') holds, then the entrepreneur's expected payoff at (X^*,ℓ^*) is equal to

(39) $\quad \left(\dfrac{1}{\beta X^*} - 1\right)\ell^*$

which is nonnegative if and only if

(40) $\quad X^* \leqq 1/\beta$.

Next we observe that for $y \neq X^*\ell^{*\beta}$, the financier can always set $R(y) = y$ to ensure that a choice (X,ℓ) with $X\ell^{\beta} \neq X^*\ell^{*\beta}$ provides the entrepreneur with a nonpositive payoff and hence is dominated by (X^*,ℓ^*) if (40) is satisfied.

Finally, in the Appendix, we show that if (38) and (40) are satisfied, then (X^*,ℓ^*) is indeed a global maximizer for the entrepreneur's second-stage problem with $y^* = X^*\ell^{*\beta}$.

In summary, for the specification (35), (36), condition (38) and (40) are both necessary <u>and</u> sufficient for the implementation of the choice (X^*,ℓ^*) through a contract $R(.)$. The financier's problem therefore takes the simple form:

$$(41)\quad \underset{X,\ell,R(.)}{\text{Max}} \quad e^{-X} R(X\ell^{\beta})$$

subject to the constraints:

$$R(X\ell^{\beta}) = X\ell^{\beta} - \frac{\ell e^{X}}{\beta X},$$

$$R(y) = y \quad \text{if} \quad y \neq X\ell^{\beta},$$

$$X \leqq 1/\beta,$$

which can be rewritten as

$$(41')\quad \underset{X,\ell}{\text{Max}} \left\{ e^{-X} X\ell^{\beta} - \ell/\beta X \right\}$$

subject to: $X \leqq 1/\beta$.

Problem (41') now has the solution

$$(42)\quad X^* = \min[1+\beta,\ 1/\beta],$$

$$(43)\quad \ell^* = \left(\beta^2 X^{*2} e^{-X^*}\right)^{\frac{1}{1-\beta}}.$$

As in previous sections, we must distinguish two cases. If

$$(44a)\quad 1 + \beta \geqq 1/\beta,$$

the financier implements the choice $(1/\beta,\ \exp(-1/\beta(1-\beta)))$, at which the entrepreneur's payoff (39) is just equal to zero. In this case, it is most profitable to push the entrepreneur to the point where he is just indifferent whether he undertakes his project or not. The financier exploits his bargaining power to the utmost. Any one who is denied a loan does not actually care because if he did get a loan he would not be better off.

In contrast, if

$$(44b)\quad 1 + \beta < 1/\beta,$$

e.g. if $\beta = 1/2$, the financier implements the choice $\left(1+\beta,\ \left(\beta^2(1+\beta)^2 e^{-(1+\beta)}\right)^{1/(1-\beta)}\right)$, at which the entrepreneur's payoff (39) is strictly positive. In this case again, consideration of the incentive effects of the contract induces the financier to refrain from extracting all the surplus that the entrepreneur's project yields. Consequently any entrepreneur who is denied finance will be strictly envious of any one who does get a contract, and we again have rationing in the sense of Stiglitz and Weiss. This result establishes our claim that under circumstances of double moral hazard, rationing may occur even if there is no constraint on the form of the finance contract that may be chosen.

To provide more insight into the nature of the solution to the financier's problem, we compare this solution to the outcome under ordinary debt or equity finance. Suppose that the financier holds a debt contract with interest obligations D as well as a share α of the entrepreneur's profits after interest payments. Then the overall contract R takes the form

(45) $$R(y) = \min(y, D) + \alpha \max(0, y - D).$$

In this case, the entrepreneur's problem (37) takes the form

(46) $$\mathrm{Max}_{X,\ell}\{e^{-X}(1-\alpha)\max(0, X\ell^{\beta} - D) - \ell\}.$$

The first order necessary conditions for a solution to problem (46) are:

(47) $$\hat{\ell} = (\beta(1-\alpha)\hat{X} e^{-\hat{X}})^{1/(1-\beta)},$$

(48) $$\hat{X} = 1 + D\hat{\ell}^{-\beta}$$

Moreover, the argument in Appendix B can be used to show that for $X^* \leq 1/\beta$, (47) and (48) are sufficient as well as necessary for a solution to problem (46).

Consider first the case $D = 0$ of pure share finance. In this case, the entrepreneur chooses $\hat{X} = 1$, the value that maximizes the product $X e^{-X}$ and that would be chosen under complete information. As in Section 3.3, the share contract involves no moral hazard with respect to project choice.

However, for $\alpha > 0$, the share contract leads to an effort level that is less than the first-best level $(\beta \hat{X} e^{-\hat{X}})^{1/(1-\beta)}$ that would be chosen under perfect information. Indeed, for $\alpha = 1$, the entrepreneur would set $\hat{\ell} = 0$ and the financier would not get anything. If he were restricted to pure share finance, the financier would therefore always set $\alpha < 1$. In this case the entrepreneur would receive a strictly positive payoff, so that the scarcity of funds would necessarily entail rationing.

In constrast, under pure debt finance ($\alpha = 0$) the effort level is equal to the first-best level for a project of risk class X. In this case there is no moral hazard with respect to effort. Instead, there is moral hazard with respect to project choice because for $D > 0$, the chosen project $\hat{X}$ necessarily involves more risk, i.e. a greater probability of failure than the first-best level.

If we compare (47) and (48) with the optimal solution (42), (43), we see that if $1/\beta \leqq 1 + \beta$, then the optimal solution can actually be implemented by a pure debt contract. For $\alpha = 0$ and

(49) $$D = \frac{1-\beta}{\beta} \; e^{-1/(1-\beta)} \; ,$$

the entrepreneur chooses $\hat{X} = 1/\beta$ and $\hat{\ell} = e^{-1/\beta(1-\beta)}$, which is exactly optimal for the financier if (44a) holds.

In contrast, a pure share contract can never be optimal for the financier because, by (42), the financier never wants to implement the first-best choice $\hat{X} = 1$. The underlying argument is quite simple: Moral hazard with respect to effort requires that the entrepreneur faces a substantial positive marginal return to effort. Under pure equity finance the marginal and the average return to effort are the same. Under debt finance, the entrepreneur's marginal return to effort will exceed his average return, because the fixed interest D has no effect at the margin. Hence under debt finance, a given incentive for effort-taking costs the financier less. To be sure, there is an adverse effect of debt finance on the entrepreneur's project choice. At the point $D = 0$ though, when the first-best project $\hat{X} = 1$ is chosen, this adverse effect is of the second order magnitude by comparison to the lowering of the cost for providing incentives for effort-taking. Therefore, $D = 0$ is never optimal.

If $1 + \beta < 1/\beta$ the optimal choice can no longer be implemented by a pure debt contract. Remarkably though, it is still possible to implement (X^*, ℓ^*) by an appropriate mixture of debt and equity finance. Specifically, if the financier sets

(50) $$\alpha = 1 - \beta - \beta^2$$

and

(51) $$D = \beta(\beta^2(1+\beta)^2 \; e^{-(1+\beta)})^{\beta/(1-\beta)} \; ,$$

the entrepreneur will choose $\hat{X} = 1 + \beta$ and $\hat{\ell} = (\beta^2(1+\beta)^2 \; e^{-(1+\beta)})^{\frac{1}{1-\beta}}$ as required by (42) and (43).

The optimal choice (42), (43) may thus be seen as a compromise. The desired project choice X^* exceeds the first-best choice, and the desired effort level ℓ^* falls short of the first-best level (even for the chosen project). For large values of β, when (44a) holds, the compromise favours debt finance, and indeed the financier pushes the entrepreneur to his individual rationality constraint. For small values of β, when (44b) holds, he prefers to both, temper the adverse effects of debt by adding some equity finance and refrain from pushing the entrepreneur to the point where he is indifferent about the project altogether.

Appendix:

Sufficiency of the First-Order-Condition Approach

In this Appendix, we show that conditions (38) and (40) are sufficient as well as necessary for the implementation of (X^*,ℓ^*) in problem (37). It will be convenient to use X and $y = X\ell^\beta$ rather than X and ℓ as arguments of the maximization. Problem (37) may then be rewritten as:

(A.1) $$\max_{X,y} v(X,y;R) \quad ,$$

where

(A.2) $$v(X,y;R) = e^{-X}[y - R(y)] - \left(\frac{y}{X}\right)^{1/\beta} \quad .$$

Along the lines of the two-stage decomposition discussed in the text, we first fix $y > 0$ and R and consider the problem:

(A.3) $$\max_{X} v(X,y:R) \quad .$$

Upon rewriting (A.2) in the form

(A.2') $$v(X,y;R) = e^{-X}\left[y - R(y) - y^{1/\beta} e^X/X^{1/\beta}\right]$$

we immediately find that for $y > 0$ the following hold:

(A.4) $$\lim_{X\to 0} v(X,y;R) = -\infty \quad ;$$

(A.5) if $v(1/\beta,y;R) < 0$, then $v(X,y;R) < 0$ for all $X \geqq 0$;

(A.6) $v(X,y;R) < 0$ for any sufficiently large X ;

(A.7) $$\lim_{X\to\infty} v(X,y;R) = 0 \ .$$

Next, we consider the first and second derivatives of v with respect to X :

(A.8) $$\frac{\partial v}{\partial X} = -e^{-X}\left[y - R(y)\right] + \frac{1}{\beta X}\left(\frac{y}{X}\right)^{1/\beta} \quad ,$$

(A.9) $$\frac{\partial^2 v}{\partial X^2} = e^{-X}\left[y - R(y)\right] - \frac{1}{\beta}\left(\frac{1}{\beta} + 1\right)\frac{1}{X^2}\left(\frac{y}{X}\right)^{1/\beta} \quad .$$

By inspection of (A.8), for $y > 0$ we find:

(A.10) $$\lim_{X\to 0} \frac{\partial v}{\partial X}(X,y;R) = \infty \ ;$$

(A.11) if $\frac{\partial v}{\partial X}\left(\frac{1+\beta}{\beta},y;R\right) \geqq 0$, then

$\frac{\partial v}{\partial X}(X,y;R) \geqq 0$ for all $X > 0$;

(A.12) $\frac{\partial v}{\partial X}(X,y;R) > 0$ for any sufficiently large X

(A.13) $\lim_{X\to\infty} \frac{\partial v}{\partial X}(X,y;R) = 0$.

Moreover, (A.2), (A.8), and (A.9) together imply for $y > 0$:

(A.14) $\frac{\partial v}{\partial X}\left(\frac{1}{\beta},y;R\right) \lesseqgtr 0$ as $v\left(\frac{1}{\beta},y;R\right) \gtreqless 0$,

and

(A.15) if $\frac{\partial v}{\partial X}(Y,y;R) = 0$ then

$\frac{\partial^2 v}{\partial X^2}(X,y;R) \gtreqless 0$ as $X \gtreqless \frac{1}{\beta} + 1$.

There are now three possibilities for the shape of the function $v(.,y;R)$:

a: If $\frac{\partial v}{\partial X}\left(\frac{1+\beta}{\beta},y;R\right) \geqq 0$, then by (A.11), (A.4) and (A.7), $v(X,y,R)$ rises monotonically from $-\infty$ to 0 as X rises from 0 to ∞ . In this case, problem (A.3) has no solution

b: If $\frac{\partial v}{\partial X}\left(\frac{1+\beta}{\beta},y;R\right) < 0$ and $v\left(\frac{1}{\beta},y,R\right) < 0$, then by (A.15), (A.4), (A.14), (A.12) and (A.7), as X rises from 0 , first $v(X,y;R)$ rises from $-\infty$ to a local maximum at some $X_1 \in \left(\frac{1}{\beta}, \frac{1}{\beta} + 1\right)$, then $v(X,y;R)$ falls to a local minimum at some $X_2 > \frac{1}{\beta} + 1$, and finally $v(X,y;R)$ rises to approach zero as X becomes large. Moreover by (A.5) and (A.7), the local maximum at X_1 is <u>not</u> a global maximum. Indeed, problem (A.3) again has no solution.

c: If $v\left(\frac{1}{\beta},y;R\right) \geqq 0$, then by (A.14), $\frac{\partial v}{\partial X}\left(\frac{1}{\beta},y;R\right) \leqq 0$. Then (A.15), (A.4), (A.12) and (A.7) imply that as X rises from 0 , first $v(X,y;R)$ rises from $-\infty$ to a local maximum at some $X_1 \leqq 1/\beta$, then $v(X,y;R)$ falls to a local minimum at some $X_2 \geqq \frac{1}{\beta} + 1$, and finally, $v(X,y;R)$ rises to approach zero as X becomes large. Now obviously $v\left(X_1,y;R\right) \geqq v\left(\frac{1}{\beta},y;R\right) \geqq 0 = \sup_{X \geqq X_2} v(X,y;R)$, so that the local maximum at X_1 actually is a global maximum. In this case, X_1 is the unique solution to problem (A.3).

The preceding discussion may be summarized in the following:

Proposition A.1: For any $y > 0$ and any contract R, X_1 is a solution to problem (A.3) if and only if $X_1 \leq 1/\beta$ and $\frac{\partial v}{\partial X}\left(X_1, y; R\right) = 0$.

Proposition A.1 immediately yields:

Corollary A.2: For any contract R, if (X_1, y) is a solution to problem (A.1), then $X_1 \leq 1/\beta$ and $\frac{\partial v}{\partial X}\left(X_1, y; R\right) = 0$.

Corollary A.3: For any $X_1 \leq 1/\beta$ and any $y > 0$, (X_1, y) solves problem (A.1) for the contract R satisfying $R(y') = y'$ for $y' \neq y$ and $R(y) = y - \frac{e^X}{\beta X}\left(\frac{y}{X}\right)^{1/\beta}$.

Upon substituting for $y = X\ell^{\beta}$, we now see that conditions (38) and (40) are indeed sufficient as well as necessary for the implementation of a choice (X^*, ℓ^*) through a contract R. Thus (41) is indeed the appropriate specification of the financier's problem.

References

Baltensperger, E. and T. Devinney (1985), "Credit Rationing Theory: A Survey and Synthesis, Zeitschrift für die gesamte Staatswissenschaft 141, 475-502.

Bester, H. (1985), "Screening vs. Rationing in Credit Markets with Imperfect Information", American Economic Review 75, 850-855.

Bester, H. (1987), "The Role of Collateral in Credit Markets with Imperfect Information", European Economic Review 31, 887-899.

Clemenz, G. (1986), Credit Markets with Asymmetric Information, Berlin, Heidelberg, New York, Springer.

Gale, D. and M. Hellwig (1985), "Incentive-Compatible Debt Contracts: The One-Period Problem", Review of Economic Studies 52, 647-663.

Hart, O. (1985), "A Comment on Stiglitz and Weiss", mimeo, Massachusetts Institute of Technology.

Hodgman, D. (1960), "Credit Risk and Credit Rationing", Quarterly Journal of Economics 74, 258-278.

Jaffee, D. and T. Russell (1976), "Imperfect Information, Uncertainty, and Credit Rationing", Quarterly Journal of Economics 90, 651-666.

Jensen, M. and W.H. Meckling (1976), "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure", Journal of Financial Economics 3, 305-360.

Keeton, W.R. (1979), Equilibrium Credit Rationing, New York and London, Garland.

Stiglitz, J. and A. Weiss (1981), "Credit Rationing in Markets with Imperfect Information", American Economic Review 71, 393-410.

The Liquidation Decision as a Principal-Agent Problem

Peter Swoboda

Summary: Substantial insights into the capital structure problem can be won if it is understood as a principal agent problem. Many principals (equity holders, debt holders, customers, tax authority etc.) employ or are dependent on an agent or a group of agents (managers). In this paper a special aspect of this comprehensive principal agent problem is analyzed: the influence of financial structure on the liquidation decision of the firm. In chapter 2 only debt holders as principals are considered. In chapter 3 a further group of principals, the customers, are introduced. Titman (1984) was the first to deal with the influence of the claims of customers on the liquidation decision. Therefore, in a first step the problem as formulated and solved by Titman (1984) is presented and it is shown that his solution is deficient. In a second step alternative solutions are considered. The concluding remarks in chapter 4 refer to the introduction of still further principals into the analysis.

Table of Contents:

1. Introduction

Following Pratt-Zeckhauser (1985, p.2), a principal agent relationship arises "whenever one individual depends on the action of another .... The individual taking the action is called the agent. The affected party is the principal". There can but need not be a contract between principal and agent protecting the interests of the principal, for example by binding the actions of the agent or defining in which way the results of the actions are divided between principal and agent (Arrow (1985, p.37)). In many cases there are no contracts. Principals then are protected by law or generally accepted rules of behavior. In most cases a combination of law and contracts is effective.

Neither law nor contracts will in general provide perfect protection

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

for principals. The reason for this is that information is not perfect and costless, and that contracting is costly. Information is perfect if all individuals assign a positive (not necessarily equal) probability to the same future states of nature. A state of nature is characterized by a deterministic outcome. Also, all individuals know all possible strategies of the agent and their outcomes in every state. With respect to economic activities perfect information also implies that all individuals know today's market prices of one unit of money income in any future state. Therefore, all individuals assign the same market value to any strategy of the agent (Franke (1981, pp. 65-68)). Under perfect information and if there were no cost of contracting, contracts would be written which prescribe exactly the state dependent actions of the manager which would maximize the market value of all firms. Therefore, contracts would also prescribe when to liquidate the firm and how to distribute the liquidation value.

Imperfect information usually implies that managers are better informed than capital owners. Ex ante, they have better knowledge on the possible states of the world and the alternatives which can be chosen (and therefore on the state dependent cash flows of the alternatives). Ex post, they are better informed on the actual state of the world and the realized alternative or profit. Because of the ex ante information gap the task of the manager now is not to fulfill a strategy agreed upon in advance but to search for an optimal strategy (Franke (1981, p. 73). Ex ante and ex post informational asymmetry, together with costly contracting, render it profitable to sign only very crude contracts between managers and capital owners. But crude contracts offer the possibility for the agent to expropriate the principal by choosing not Pareto efficient actions.

Gale-Hellwig (1985) haven shown that, if there is only ex post informational asymmetry and if capital owners can trigger off costless bankruptcy, pure debt contracts are optimal and avoid agency costs: Whenever the manager maintains that he cannot repay the debt, the debt holders can bring about bankruptcy and can produce perfect information without cost. In the other states the information asymmetry would not hurt the debt holders. Including ex ante informational asymmetry, pure debt contracts need not be optimal. Nevertheless we will concentrate on pure debt contracts, possibly enriched by information and decision rights of the debt holders, because they are, together with equity financing, dominant in real world.

There is a vast amount of papers on incentives for managers to expropriate capital owners and how capital owners tend to protect themselves. In this paper only the incentives to decide inefficiently between liquidation (bankruptcy) and the ongoing (reorganization) of the firm is analyzed. We assume that firms are run by owner-managers or that managers act in the interest of owners so that the agency costs of equity financing can be neglected. In the next chapter we analyze the influence of debt capital on the liquidation decision, in chapter 3 the impact of claims of customers.

2. The Liquidation Decision with Creditors as Principals

A wholly equity financed firm should be liquidated if the *liquidation value*, the revenue gained by selling the assets of the firm, exceeds the *value of the ongoing firm* (Haugen-Senbet (1978)). Debt financing may distort this decision. The higher the debt level the lower is the share of equity holders in the case of liquidation. Usually debt exceeds the liquidation value whenever liquidation is considered. Then the share of equity holders is zero. Consequently equity holders loose nothing or very little by continuing the firm. But they can realize considerable gains, on the one side by drawing salaries or withdrawing cash from the firm, on the other side merely by the chance that the investment program will have a favorable outcome and liquidation can be avoided in the long run. The latter aspect can be shown more formally. It is assumed:

$$D > L > V$$

D = nominal debt
L = liquidation value
V = (expected) value of the ongoing firm

$$V = p_1V_1 + p_2V_2$$
$$V_1 > D, \quad V_2 < D$$

p_1, p_2 = probabilities
V_1, $V_2 = L_2$ = possible values of the ongoing firm

If V_1 is realized the firm is saved. If V_2 is realized the firm is in still greater distress and is liquidated (= L_2).

The values of the equity position and the debt position in the case of liquidation and continuation are:

VE = value of the equity position
VD = value of the debt position

VE (in the case of liquidation) = 0
VE (in the case of continuation) = $p_1(V_1 - D) > 0$
VD (in the case of liquidation) = L
VD (in the case of continuation) = $p_1D + p_2V_2 < V < L$

Hence the equity holders will be in favor of continuation, although the liquidation value is higher than the value of the ongoing firm. In addition to that there is an incentive to increase the riskiness of the investment programm. Even projects with negative present values can be profitable for the agent if $p_1(V_1 - D)$ is increased. But if an inefficient project is realized and the absolute share of the equity owners is increased the share of the debt holders must decrease.

How can an inefficient decision between liquidation and continuation of the firm be avoided or improved?

a) Debt holders grant only secured credits. They do not accept risk. Then they cease tc be principals - they cannot be affected by the owner-manager. This alternative would restrict severely the debt level of firms and would cause transaction costs for the creation and control of security rights. Reality seems to show that this alternative is not advantageous. Debt financing seems to have advantages which outweigh its agency costs.

b) Debt holders grant secured and unsecured credit. There is a priority order for the unsecured loans. The investment strategy is agreed upon in advance or creditors may influence investment decisions. Obviously, the transaction costs of this alternative would be very high.

c) As alternative b). But in addition to the contractual agreements between principal and agent, the principals (debt holders) are protected by a bankruptcy act. The purpose of bankruptcy regulations is threefold. First, penalties or troubles for the managers, connected with a bankruptcy, may prevent them from certain actions. Second, bankruptcy transfers the decision power and/or property rights from the equity holders to the debt holders, i.e. from the group which has lost most or all of its property to the group which has still some capital at stake. By this transfer further agency costs should be diminished. Third, the bankruptcy procedure prevents debt holders from fighting one against another and thereby helps to avoid transaction costs (Drukarczyk (1983)).

In the following we will concentrate on alternative c) and discuss some problems of bankruptcy regulations. The most crucial issues of a bankruptcy act are the conditions under which bankruptcy can be brought about und what rights should be given to the debt holders and left with the equity owners. These questions cannot dealt with in detail. Only some important points shall be stressed.

First, it is not sufficient to give to debt holders the right to decide on the future strategy of the firm without changing their and the equity owners' claims. In this case the debt holders would be interested to deviate from the overall optimal strategy, in a way exactly opposite to the incentives of the owner-manager. This can be shown as follows. Assume now that

$$D > V > L$$

so that the operations of the firm should be continued. Again

$$V = p_1V_1 + p_2V_2$$

$$V_1 > D, \quad V_2 < D$$

hold, so that the value of the ongoing firm for the debt holders is

$$VD = p_1D + p_2V_2 < V.$$

Since VD can be easily smaller than L, the debt holders can prefer liquidation, although the ongoing of the firm is efficient. The debt holders may also turn to a less risky, inefficient strategy which is characterized by a lower V_1 and a higher V_2 and therefore higher VD, if p_1 and p_2 are unchanged.

To eliminate agency problems in bankruptcy it is therefore necessary to grant to debt holders also the right to change their and the equity owners' claims. For example, the bankruptcy act could allow the debt holders to opt to become equity holders themselves, or to sell the firm to a trouble shooter, or to sell the firm to the original equity owners for a part of their claims. They will prefer the latter alternative if they think that the manager-owner in charge up to bankruptcy is the better manager and can pay more than a trouble shooter could. All this implies that the debt holders turn from principals into agents, at least as far as the decision between liquidation and the ongoing of the firm is concerned. In the case of a settlement with the equity owners the debt holders change again their role and they become principals once more.

Second, if the debt holders decide in favor of a settlement with the owner-manager, the agency costs of debt capital will be reduced but not eliminated. Because of the better information of the manager-owner and his incentive to continue the operations of the firm whether it is efficient or not, the debt holders cannot be sure that their reduced claims will be fulfilled and that the equity owners will not ravish the firm. The manager qualities of the owner must be quite high that to let him in charge is the best alternative for the debt holders! Even if debt holders are paid out in cash they do not know the price limit the owner-manager is prepared to accept. Franke (1981, p.79) points out that the more generously debt holders behave, the worse their bargaining position is in future game situations. This external effect diminishes further their interest to reach a settlement with the equity owners.

Third, the structure of the claims within the group of debt holders may be quite different. Debt holders with security rights may be in favor of liquidation if the quality of their security rights decreases with time. Other debt holders may be in favor of continuing the operations. Or suppliers, in opposition to other debt holders, may prefer the continuation of the firm because of the profits they expect from future agreements. Side payments may be used to eliminate these effects.

Fourth, the third aspect is aggravated if debt holders have heterogeneous expectations on the liquidation value and the value of the firm as an ongoing entity. In this case the rules for deciding between liquidation and continuation become crucial (Franke (1986)).

3. The Liquidation Decision with Creditors and Customers as Principals

In this chapter a second group of principals, the customers, is introduced. The firm grants to its customers guarantee rights and/or the right to require maintenance at low cost. The customers loose these claims in the case of liquidation. This creates an incentive to liquidate the firm in states where the value of the ongoing firm is still higher than the liquidation value. Titman (1984) was the first to analyze this problem in detail. In a first step the problem as formulated and solved by Titman is presented and it is shown that his solution is deficient. In the second step we will search for a preferable solution.

3.1 The Problem and the Solution of Titman

Titman (1984) starts with a single agent - single principal problem. The agent is the owner-manager, the principal is the group of customers. The firm is wholly equity financed. It produces machines. The "production of machines and their maintenance (e.g., spare parts) exhibits joint economies of scale (Titman (1984, p.140/141)). Maintenance is sold for its marginal cost which is lower than the cost for maintenance provided by a third party. Therefore, liquidating the firm would cause losses to customers, amounting to the additional costs to be paid to a third party which provides maintenance at higher cost. The owner-manager will liquidate the firm if

$$L > V_o.$$

V_o = value of the firm for the capital owners

In the interest of both, capital owners and customers, however, the firm should only be liquidated if

$$L > V_o + C = V \quad \text{or} \quad V_o < L - C.$$

C = additional cost for customers when employing outside suppliers of maintenance

Titman (1984) solves the problem by introducing a *second and third principal* - debt holders and holders of preferred stock. The second and third principal control the agent in the interest of the first principal. The most simple model functions as follows (Titman (1984, p. 148)): The firm issues short term debt (D) with a maturity value equal to L - C, and preferred stock (P) in an amount $P > C$. Then the debt holders would initiate bankruptcy (liquidation) whenever

$$D = L - C = V_o,$$

which is the correct criterion.On the other side, the equity holders are no longer interested in earlier liquidation. If they liquidate whenever

$$L = V_o,$$

L would be divided between debt holders and the holders of preferred shares. The share of equity holders would be zero.

The main shortcoming of Titman's (1984) solution is the implicit assumption that the capital owners will not form a coalition. If they do it is straightforward to see that it is in their interest to liquidate if

$$L = V_o > L - C,$$

sharing the difference between L and L - C among them in a way so that every capital owner prefers liquidation to the ongoing of the firm.
A further shortcoming is that the solution depends on L being certain at the time when debt is raised.

3.2 Can the Problem be Solved by Changing the Price Policy of the Firm?

As Titman (1984, p.140/141) points out the problem arises because production of machines and maintenance exhibit joint economies of scale, and because maintenance is sold for its marginal cost. If this is so, why can't the problem be eliminated by increasing the price for maintenance to a level a third party would charge, and, for exchange, by lowering the price for the machines? If this could be done the customers would not be damaged by the liquidation decision. They have already received the full advantage from the joint production via a lower price of the machine. Unfortunately it can be shown that such a policy is not sustainable in a competitive market.

The following notation is used:

A = long term marginal cost of producing the machine

R(t) = long term marginal cost of maintenance for a machine which is t years old

i = rate of discount

n = life of the machine

Ann(n) = $i/(1 - e^{-in})$ = annuity factor for n

The *assumptions* are:

a) The optimal life of the machine is the period over which the uniform equivalent annual cost is minimized. We therefore assume that the machines are reinvested by identical plants until infinity. This assumption is not important. The same result can be derived assuming that the machine is not reinvested so that the optimal life is the period over which the present value is maximized.

b) R(t) is a strictly increasing function of t, the age of the machine. All other costs and receipts from using the machine are not dependent on the age of the equipment and therefore will not influence the life of the investment. R(t) eventually will exceed the uniform equivalent annual cost of the machine. These assumptions - together with assumption c) - are sufficient for a finite optimal life of the machine.

c) The salvage value of the machine is zero.

d) The market for machines is competitive. Assumption d) implies that the producer may charge A and R(t) to his customers without being eliminated from the market. Also, since A and R(t) include interest on capital and all other long term marginal cost, the producer is interested to maintain production if he receives A and R(t).

Proposition: Given the assumptions a) - d), the prices A and R(t) are optimal in the sense that any other price policy A' and R'(t), with A' < A and R'(t) > R(t) either causes losses for the producer or higher costs for the users so that the machine either would not be produced or not bought.

This proposition is also valid for A' > A and R'(t) < R(t). However, since such a policy would even enhance the agency problem it is not of interest.

Proof: 1) Given the prices A and R(t) the optimal life n* of the machine for the user is the life which minimizes the uniform equivalent annual cost:

$$\text{Min: } \left(A + \int_0^n R(t)e^{-it}dt\right)\text{Ann}(n). \qquad (1)$$

The first order condition is given by:

$$\left(A + \int_0^{n^*} R(t)e^{-it}dt\right)\text{Ann}(n^*) = R(n^*). \qquad (2)$$

2) If the prices for maintenance are increased to R'(t) > R(t) for all t, and the price of a machine is reduced to A', the new optimal life of the investment n' is given by:

$$\left(A' + \int_0^{n'} R'(t)e^{-it}dt\right)\text{Ann}(n') = R'(n'). \qquad (3)$$

R'(n') can be greater or smaller than R(n*), since n' need not be equal to n*.

a) If R'(n') > R(n*), also the uniform equivalent annual cost given the policy (A', R'(t)), the left hand side of (3), must be higher than the annual cost of the policy (A, R(t)):

$$\left(A' + \int_0^{n'} R'(t)e^{-it}dt\right)\text{Ann}(n') = R'(n') > R(n^*) =$$
$$\left(A + \int_0^{n^*} R(t)e^{-it}dt\right)\text{Ann}(n^*). \qquad (4)$$

Therefore, the policy (A', R'(t)) is not competitive.

b) If, however, R'(n') ≤ R(n*), the policy (A', R'(t)) implies a loss for the producer. R'(n') ≤ R(n*) is, because of assumption b), only compatible with n' < n*. Therefore,

$$\left(A + \int_0^{n'} R(t)e^{-it}dt\right)\text{Ann}(n') > R(n^*) \geq R'(n') =$$
$$\left(A' + \int_0^{n'} R'(t)e^{-it}dt\right)\text{Ann}(n'). \qquad (5)$$

The first inequality holds because n is not optimal with respect to A and R(t). Inequality (5) shows that the costs of the producer of the machine over the life of n' (first part of (5)) is higher than his income (last part of (5)); q.e.d.

3.3 An Alternative Solution

Changing the price policy, therefore, provides no solution to the agency problem. There is an alternative solution, however. The firm could accept the obligation to put C, the claims of the customers, into a reserve account. This obligation must have *first priority* if the firm is liquidated. Furthermore, the customers must have the right to trigger off bankruptcy if the liquidation value of the company is reduced to C and to take over the firm.

Such a solution which could also be demanded by law would guarantee that owners and debt holders are only interested to liquidate if $L - C > V_o$. It is free from the danger that owners and debt holders form a coalition to expropriate customers. Since the customers have first priority in the case of liquidation, their losses, in any state of the world, are clearly equal to or smaller than those in the solution of Titman (1984). This, however, is not important in itself, since this effect would be offset in the prices of the products.

4. Some Remarks with Respect to further Principals

There may be several further principals who hold claims which are similar to those of the customers and which could be responsible for an inefficient liquidation decision. Such principals include the tax authority and the social security administration. Common to these principals is that they do not decide *directly* to grant credit to the firm. For example, the claim of the social security administration arises by employing workers. These principals cannot react to a deterioration of the financial standing of the firm by demanding security rights for claims which will arise in future. Therefore, in accordance with the solution found for customers, it can be argued that priority should be given to these compulsory debt holders as well. Much the same is true with respect to the salaries of the firm. Although workers can leave the firm if its standing is deteriorating they must incur high transaction costs. They also can be expropriated by the liquidation of the company if they do not have priority. This is the reason why the tendency in bankruptcy law reforms to abolish the priority rights of these groups of principals does not seem to be well founded.

References

Arrow, K.J. (1985): The Economics of Agency, in: Pratt, J.W.; Zeckhauser, R.J. (Eds.): Principals and Agents: The Structure of Business, Boston, 37-51

Drukarczyk, J. (1983): Kreditverträge, Mobiliarsicherheiten und Vorschläge zu ihrer Reform im Konkursrecht, Zeitschrift für Betriebswirtschaft, 53, 328-349

Franke, G. (1981): Information, Property Rights, and the Theory of Corporate Finance, in: Derkinderen, F.J.; Crum, R.L. (Eds.): Readings in Strategy for Corporate Investment, Boston, 63-83

Franke, G. (1986): Zur Festlegung von Abstimmungsregeln im Insolvenzverfahren, Zeitschrift für Betriebswirtschaft, 56, 614-630

Gale, D.; Hellwig, M. (1985): Incentive Compatible Debt Contracts: The One-Period Problem, Review of Economic Studies, 52, 647-663

Haugen, R.A.; Senbet, L.W. (1978): The Insignificance of Bankruptcy Costs to the Theory of Optimal Capital Structure, Journal of Finance, 33, 383-393

Hax, H.; Marschdorf, H.J. (1983): Anforderungen an ein Insolvenzrecht aus betriebswirtschaftlicher Sicht, Betriebswirtschaftliche Forschung und Praxis, 35, 112-130

Pratt, J.W.; Zeckhauser, R.J. (1985): Principals and Agents: An Overview, in: Pratt, J.W.; Zeckhauser, R.J. (Eds.): Principals and Agents: The Structure of Business, Boston, 1-35

Swoboda, P. (1987): Kapitalmarkt und Unternehmensfinanzierung - Zur Kapitalstruktur der Unternehmung, paper in print, Graz

Titman, S. (1984): The Effect of Capital Structure on a Firm's Liquidation Decision, Journal of Financial Economics, 13, 137-151

On Stakeholders' Unanimity

Jochen E. M. Wilhelm

Summary: This paper provides an approach to quantify the concerns of the different parties having stakes in a business activity, in terms of the implied consumption opportunities; it outlines inherent potentials for conflict and presents possible strategies to resolve the complex problem. We show that the market value of the initial endowment with financial assets and labour contracts may be regarded as an indicator that allows for preference independent ordering of the desirability of implied consumption opportunity sets under particular assumptions on how financial markets work and whicn allocative features they have. Thus the market value of the inital endowment can be used as objective function serving as the basis for a quantitative analysis of the wellknown conflicts between owners and creditors, and managers and owners, respectively. This analysis particularly reveals the importance of an appropriate designed wage scheme for harmonizing the interests of managers and owners. A fully satisfactory solution for all the parties concerned seems however difficult to design because of a manager's incentive to impair both owners and creditors.

Table of Contents:

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

1. Introduction

Having impact on the welfare of many people or groups of people, respectively, entrepreneurial decisions are, in turn, driven by many people or groups of people, respectively. Usually, at least some of those being affected have neither legal nor actual possibilities to influence on or to participate in decision making and executing. People who are able to influence business decisions and their execution can be found among the following groups (enumerated according to their relative strength of influence): Managers, owners, financiers, suppliers, clients, insurers, workers (there may apply a reversed order when accounting for their participatory rights), competitors and 'public opinion'. In many cases people are exposed to the dispositions of decision makers voluntarily and, thus, are able to terminate their exposition any time it seems advisable to them. Those cases are referred to as 'principal-agent-relationships'. In order to get the agent to take action according to the principal's concerns, or at least to take them into consideration to some extent, requires the principal's awareness of what his/her concerns really are. At the same time, a principal being anxious about the agent's willingness to take action according to his/her concerns must make up his/her mind about the agent's interests. The nature and size of the divergence between their interests motivate the nature and rigour of contractual regulations between principals and agents, i. e. the agent's competences and his/her compensation pattern, the principal's control mechanisms etc.

As we have outlayed above, coping with corporate policy requires not a single principal but a whole lot of them to be regarded thus posing the question to the decision maker whose concerns must not and whose concerns may be neglected. The institution of private ownership suggests that there should be taken care only about concerns which have, when touched, in turn an impact on the owners' sphere of interest. In accordance, a financier's concerns should be taken into consideration only to the extent that the financier might otherwise restrict the decision maker's opportunity set on whatever legal, contractual, or actual reasons, or might influence the outcomes of certain measures. The stakes of other non-owners should be treated in the same spirit; the decision maker only has to figure out their interests as far as their responses to the firm's decisions have a direct or indirect impact on the owners' interests.

Anticipating these indirect effects, however, requires to assess first, in as much different stakeholders are concerned about the firm's decisions. It takes a second step to analyse how people are likely to react and, thereby, to cause changes in the opportunity set or in results. The focus of this paper will be on the first step.

Our analysis of the effects of certain entrepreneurial decisions on the stakeholders' interests will, actually, rely on two restrictive assumptions which allow for a rigorous treatment: firstly we assume that the concern of each affected party may be represented by the consumption opportunities related to its stakes thereby taking account for the fact that different parties are affected in different ways. Secondly, we restrict consumption opportunities to result from no more than those three sources which are, mainly and unambiguously, related to a firm's fate, namely stocks, bonds or credit contracts, respectively, and labour contracts.

At first glance, individual concerns seem to be, at least to some extent, determined by unobservable personal preferences. Furthermore, it should be recognized that the impact of a particular firm's actions on its stakeholders' consumption opportunities may be weakened or fully absorbed by properly adjusted economic transactions done by those stakeholders on their own account. Thus, integrating the stakeholders' economic environment into the analysis seems natural, if not necessary. Fortunately, integrating the economic environment by introducing appropriate conditions proves as a vehicle to eliminate the influence of personal preferences on the individual assessment of particular decisions rather than as an additional complication. The appropriate methodology is provided by the 'Arbitrage-Theory' in combination with certain assumptions about the allocative properties of financial markets. In this respect our paper may be looked upon as an extension of DeANGELO's (1981) and WILHELM's (1983) work (the machinery developed by MAKOWSKI would do as well (see MAKOWSKI/PEPALL (1985)).

In the following paragraph we will give a sketch of our technique of analysis. In subsequent sections we will at first present a proposal to deal with the conflict between owners and financiers, on the one hand, and with the conflict between owners and managers on the other hand, finally coming up with a reconciliation of these separate analyses.

2. The Dependence of Stakeholders' Consumption Opportunities on Corporate Policy

2.1 Notations and Conventions for Model Representation

We assume that there are only two points in time $\tau \in \{0,1\}$, consumption in τ is represented by cash-outflow c_τ. $\omega \in \Omega$ indicates a particular state of all possible states of the world in $\tau = 1$; X_Ω denotes the set of all random variables on Ω. The set of (legally) permissible entrepreneurial decisions in $\tau = 0$ is referred to as A. Deciding for a particular action $a \in A$ has the following effects:

- net payment of $z_o(a)$ in $\tau = 0$ to the *owners* ($z_o(a) < 0$ means that owners are required to supply new equity)
- payment of $z_1(a,\omega)$ in $\tau = 1$ to *all financiers*, owners or not, as a lump sum payed in accordance to the proper sequence of satisfaction
- payment of $h_o(a)$ and $h_1(a,\omega)$ to the *individual* in question - considered as an employee - as wage income.

The following notations are also used throughout the paper:

α indicates an *individual's* share in total equity in the initial state (initial endowment), α^n the respective share after trading has taken place

β indicates an *individual's* share in total debt in the initial state (initial endowment), β^n the respective share after trading has taken place

y represents the *individual's* portfolio of all other tradable securities in the initial state (initial endowment), y^n the respective portfolio after trading has taken place

x indicates the *individual's* initial position in securities that yield the sure rate of interest $r = R - 1$ (initial endowment), x^n the respective position after trading has taken place

Z is the vector of uncertain cash flows resulting from each of all other tradable securities in $\tau = 1$ (Z is a m-dimensional random vector: $Z \in X_\Omega^m$)

P is the vector of these securities' current prices in $\tau = 0$ (P is an ordinary m-dimensional vector)

p_E denotes the current ex-dividend market value of equity of the firm under consideration in $\tau = 0$

D indicates the face value of debt to be paid back to the firm's creditors in $\tau = 1$

p_D denotes the current market value of the firm's debt in $\tau = 0$

The following assumption helps to simplify a rigorous representation of an individual's interests and his/her exposition to corporate policy:

<u>Assumption 1</u>: Individuals' only concern is consumption c_o in $\tau = 0$ and consumption c_1 in $\tau = 1$. At each point in time they prefer more consumption to less.

Using the definitions given thus far, any individual - with all his/her shares in equity and debt of the respective firm to his/her unrestricted disposition but given no choice to switch or terminate his employment - will face the following *consumption opportunity set*:
$C(a;\alpha,\beta,y,x) =$

$$= \{(c_o(a;\alpha^n,\beta^n,y^n,x^n) \;,\; c_1(a;\alpha^n,\beta^n,y^n)) \in \mathbb{R} \times X_\Omega \mid (1), (2) \text{ und } (3)\},$$

where

$$c_o(a;\alpha^n,\beta^n,y^n,x^n) = h_o + \alpha^n z_o + (x - x^n) + P \cdot (y - y^n) + \qquad (1)$$
$$+ (\alpha - \alpha^n)(z_o + p_E) + (\beta - \beta^n)p_D$$

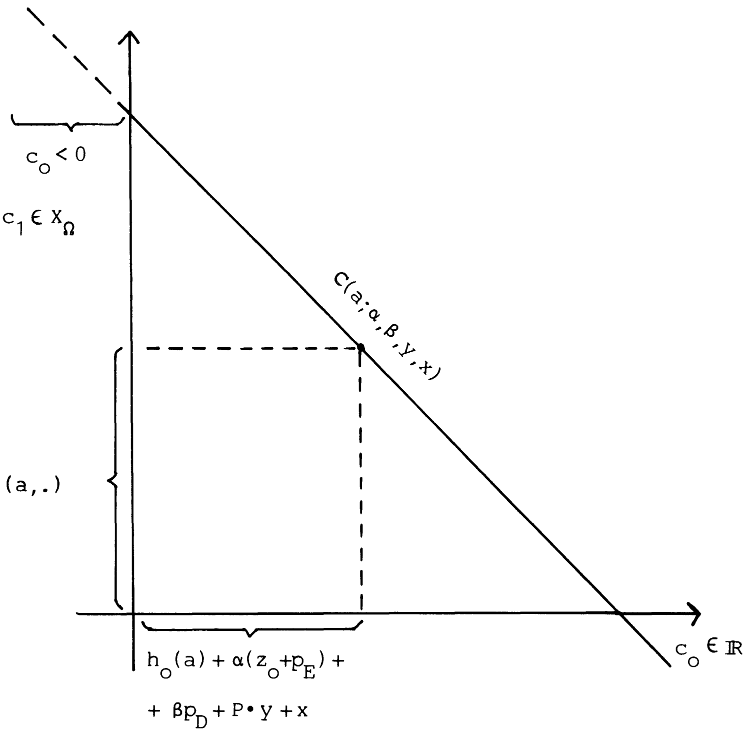
$$c_1(a,\omega;\alpha^n,\beta^n,y^n,x^n) = h_1 + \alpha^n \max\{z_1 - D, 0\} + \beta^n \min\{D, z_1\} + \qquad (2)$$
$$+ Zy^n + R \cdot x^n$$

$$\alpha^n,\beta^n,x^n \in \mathbb{R} \;,\; y^n \in \mathbb{R}^m \;;\; c_o(a;\alpha^n,\beta^n,y^n,x^n) \geq 0 \qquad (3)$$

The superscript ^{n} throughout indicates the respective decision variables' values after trading has taken place. We have omitted explicitly stating a variable's dependence on an action a or the state of the world ω, respectively; this has been done since a detailed discussion of this point will follow subsequently.

Please refer to Appendix I for a rigorous representation of the mathematical structure of the consumption opportunity set. Figure 1 provides a graphical representation:

Figure 1: Consumption opportunity set with fixed initial endowment



The consumption opportunity set $C(a;\alpha,\beta,y,x)$ consists of all those combinations of (certain) consumption in $\tau = 0$ and (uncertain) consumption in $\tau = 1$ that may be generated from (a) the proceeds from selling securities, from (b) cash-flows resulting from capital as well as (c) wage income given the initial endowment α,β,y,x. In (1) p_E indicates the ex-dividend market price of equity (the actual price is cum dividend) whereas p_D indicates the market price of debt. In (2) $\max\{z_1 - D, 0\}$ reflects the owners' limited liability whereas $\min\{D, z_1\}$ reflects the *par conditio creditorum* of debt capital.

According to (1) and (2) the consumption opportunity set (i. e. the individual's interests are) in full generality at least potentially affected by action $a \in A$ via:

- wage income ($h_o(a)$; $h_1(a,\omega)$)
  - -- as directly related to performance (contingent fee) or
  - -- as indirectly related to performance (for example via loss of employment in case of insolvency)
- the firm's payments to the owners and creditors ($z_o(a)$, $\max\{z_1(a,\omega) - D, 0\}$, $\min\{z_1(a,\omega)$, $D\}$)
- market prices of the firm's equity and debt ($p_E(a)$, $p_D(a)$)
- payments resulting from the set of all other tradable securities ($Z(a,\omega)$)
- market prices of these securities ($P(a)$)
- the riskless rate of interest ($R(a)$)

The possible dependencies are determined by the respective individual's and the market's perception of what the particular action might be that has been taken.

2.2 Simplifying Assumptions and Their Consequences

The following assumption contributes to a less complex approach by suspending the necessity of an explicit consideration of consumption and risk preferences (cf. WILHELM (1985)).

Assumption 2: There is an arbitrage equilibrium before and after the decision $a \in A$ has been made; the resulting positive linear price functional (ROSS (1978), HARRISON/KREPS (1979)) before decision making is denoted by π, that after decision making by π^a.

Assumption 2 implies:

$$p_E = \pi^a(\max\{z_1(a,.) - D, 0\})$$

$$p_D = \pi^a(\min\{z_1(a,.), D\})$$

$$P = \pi^a(Z)$$

$$\frac{1}{R} = \pi^a(1)$$

By $z_1(a,.)$ is meant the whole random variable z_1 given action $a \in A$; similar symbolics will be used from time to time.

Assumption 3: ('competitivity'): $\pi = \pi^a$; i. e. the market's method of how to transform securities' risk/chance characteristics into market prices is not affected by the respective firm's actions.

Assumption 3 directly implies that the rate of return on riskless securities remains unaffected by the firm's decision and that the market prices of all other tradable securities are affected only to the extent that the payments they provide at $\tau = 1$ are improved or vitiated by the action in question. At least three situations might be imagined that potentially imply such a dependency:

- competitors might profit or suffer from a particular action of the respective firm
- cooperating firms might benefit from synergy effects
- contingent claims (contingent on the respective firm's equity or debt) might be contained in the set of tradable assets (for example options, warrants, convertible bonds etc.).

Such dependences will be excluded by the following:

Assumption 4: The cash-flows of all other tradable securities are independent of the action to be taken.

The assumption has implications that will be discussed below in connection with Assumption 5.

With regard to Assumption 3, Assumption 4 directly leads to the conclusion that the market prices $P = \pi^a(Z) = \pi(Z)$ are independent of the action to be taken.

Another fundamental assumption is as follows:

Assumption 5: ('spanning') The cash-flows associated to shares in the respective firm's equity or debt may be reproduced by appropriate combinations of the other (tradable) securities, i. e. there are portfolio y_E^a and y_D^a with

$$\max\{z_1(a,.) - D,0\} = Z \cdot y_E^a$$

and

$$\min\{z_1(a,.) , D\} = Z \cdot y_D^a ,$$

respectively.

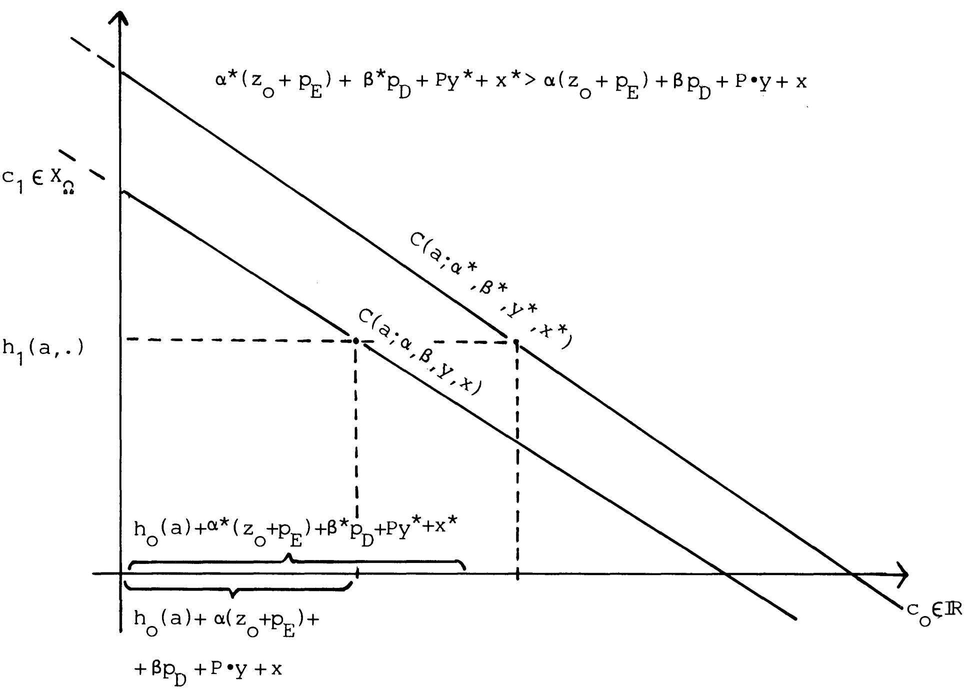
A critical implication of Assumption 5 concerning the availability of information should actually be recognized; the fact that anybody who wants to reproduce his stake in the respective firm by a portfolio of other securities *must be aware of* the action $a \in A$ to be taken *in advance*. This assumption should be perceived as being fundamental with respect to the separation result which we aim at: any investor who wants to get independent of corporate policy with respect to his consumption and risk preferences must know this policy in advance, otherwise he won't be able to set up the needed hedging portfolio (a similar argument applies within the approach of MAKOWSKI which does *not* presume spanning (cf. MAKOWSKI/PEPALL (1985); there it is shown that individuals can mimic their personal optimization problem by an optimization problem whose constraints are market determinded under competitivity; but in order to actually realize the optimal position they must know in advance, how this position can be achieved through actually traded market instruments; this seems to be tantamount to the situation under spanning).

We are now able to formulate our first separation result:

Separation I: Given Assumption 2 to 5 consumption opportunities do not depend on the respective firm's action except for wage income and the market value of the initial endowment with financial assets (they do *not* depend on the structure of this endowment). C.p., a higher market price is equivalent with a preferred consumption opportunity set, preferred according to Assumption 1.

Proof: A proof is given in Appendix II; Figure 2 shows a simplified graphical representation of the situation.

Figure 2: Consumption opportunity set corresponding to a varying initial endowment with securities



A problem still waiting for its solution is the possible dependence of the consumption opportunity set on the structure of wage income (see MAYERS (1972), BRITO (1977) on portfolio construction in presence of non-marketable income). But again, the problem may be solved with the aid of a kind of 'spanning'-assumption:

Assumption 6: ('spanning' of wage income) Wage income in $\tau = 1$ may be duplicated by a portfolio y_A^a of other traded securities, i.e. there is a portfolio y_A^a satisfying

$$h_1(a,.) = Z \cdot y_A^a$$

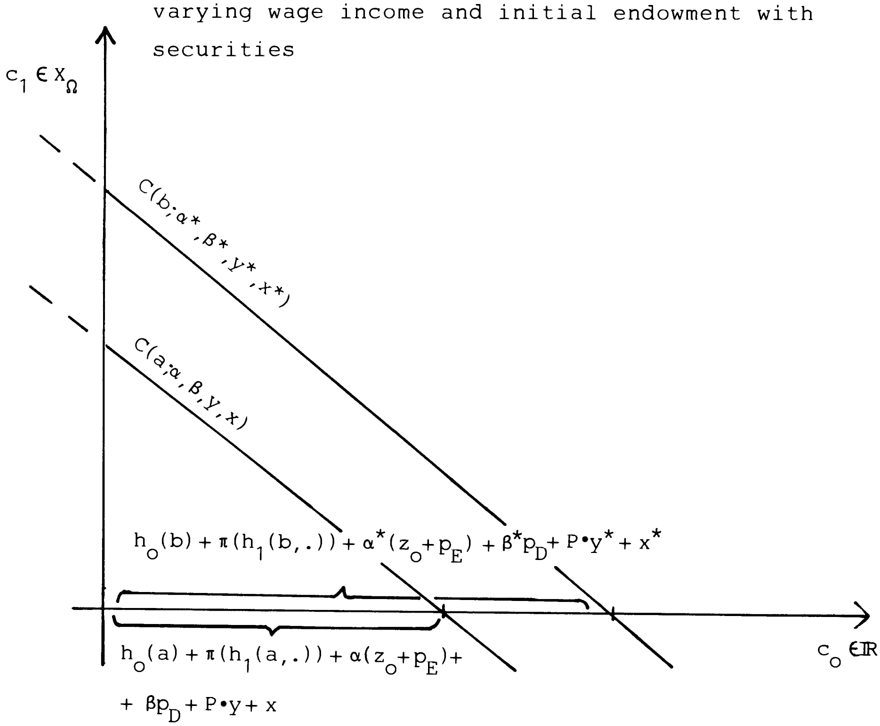
Assumption 6 requires a degree of availability of information comparable to that of Assumption 5: eliminating repercussions of corporate policy on one's own sphere of interests in terms of individual consumption and risk preferences, necessitates the knowledge of the policy to be taken; it is this knowledge which enables the person in question to set up an appropriate hedging portfolio.

Furthermore Assumption 6 implies the following separation result:

Separation II: Given Assumption 2 to 6 consumption opportunities do not depend on the considered firm's action except for the market value of wage income and of the initial endowment with financial assets (their composition and risk structure do *not* matter). A higher market price corresponds to preferred consumption possibilities, preferred according to Assumption 1.

Proof: A formal proof is given in Appendix III; Figure 3 gives a simplified graphical representation.

Figure 3: Consumption opportunity set corresponding to a varying wage income and initial endowment with securities



Separation II enables us to establish a preference-free objective function attributable to each shareholder according to his/her initial endowment; of course, this is true only for individuals whose interests are solely based on shares in equity and/or debt and labour contracts, respectively. The objective is:

> Maximize the sum of market values of wage income and of the initial endowment in shares in the firm's equity and debt:

$$h_o(a) + \pi(h_1(a,.)) + \alpha(z_o(a) + p_E(a)) + \beta p_D(a) \longrightarrow \max !$$

This objective function will be our basic tool as we proceed with the analysis.

3. Two Typical Sources of Conflict and Their Interaction

3.1 The Conflict Between Owners and Creditors

A conflict between owners and creditors arises whenever, at a given level of debt obligations, a decision is to be made about the optimal use of funds - already raised or not - . This will be demonstrated by the following consideration:

Letting q indicate the positive stochastic discount factor transforming uncertain cash flows into market values (the existence of q is a consequence of arbitrage equilibrium; cf. WILHELM (1985)) we have:

$$p_E(a) = \int_0^{\infty}\int_D^{\infty}(z-D)qf(z,q,a)dzdq$$

for the ex-dividend market value of equity, and

$$p_D(a) = D\int_0^{\infty}\int_D^{\infty} qf(z,q,a)dzdq + \int_0^{\infty}\int_0^{D} zqf(z,q,a)dzdq,$$

for the market value of debt, thereby f(z,q,a) indicating the common density function of cash flow z_1 and discount factor q given action $a \in A$.

$$\frac{d}{da}\left[\alpha(z_o(a)+p_E(a))+\beta p_D(a)\right]=0$$

and

$$\frac{d}{da}\left[\alpha(z_o(a)+p_E(a)+p_D(a))+(\beta-\alpha)p_D(a)\right]=0,$$

respectively, entail an optimum when not taking account for wage income.

Expression $p_G(a) = z_o(a) + p_E(a) + p_D(a)$ represents the firm's total market value; obviously, maximization of the market value of a firm produces a solution to the above stated *individual* maximization problem if the individual in question holds equal shares in equity and debt ('proportional financiers'). In general, the following (necessary) condition for an optimum holds:

$$\alpha\frac{d}{da}p_G(a)+(\beta-\alpha)\left[D\cdot\frac{d}{da}\int_0^{\infty}\int_D^{\infty}qf(z,q,a)dzdq+\int_0^{\infty}\int_0^{D}zq\frac{\partial f}{\partial a}(z,q,a)dzdq\right]=0$$

The above condition implies that it is sufficient for maximization of a firm's total value to be an optimal strategy for any stakeholder - regardless of how his/her stake looks like - that at the maximum of the respective firm's total market value the probability of solvency and the density function within the range of insolvency do not depend on the decision. These conditions are, of course, met whenever default is completely unlikely.

Except for the two cases just mentioned, i.e. that of 'proportional financiers' and of default free business, a conflict of interest arises whenever enhancing a firm's total market value inevitably reduces the market value of its debt, simultaneously.

A conflict does not arise, however, if the firm's action is decided upon and honestly announced to the market before additional debt is incurred: in that case the supplied funds are directly paid to the owners resulting in total payments $z_o(a) + p_D(a)$ which amounts to the following market price of *equity*:

$$z_o(a) + p_D(a) + p_E(a) = p_G(a);$$

as $\beta = 0$ (there is no initial endowment with debt) maximization of the firm's total value $p_G(a)$ is unanimously supported.

The following measures appear to be proper means to limit potential conflicts of interest between owners and creditors (cf. a. o. SMITH/ WARNER (1979)):

- the instrument of proportional financiers
- measures guaranteeing financiers a far reaching protection against default via
  - -- legal protection provided by accounting rules
  - -- collateral
  - -- agreement on and surveillance of negative clauses
  - -- shortening credit terms.

All but the first possibility that actually comes up with an elimination of borrowing as a separate means of raising funds, involve some danger of suboptimal decisions with respect to aggregate welfare: there is no guarantee that maximization of the firm's total market value is compatible with (perfect) default protection (see HAUGEN/ SENBET (1981) and KUDLA (1984) for the use of options or warrants to eliminate incentives to impair creditors).

3.2 The Conflict Between Owners and Managers

A seminal paper on the issue of owner-manager-conflicts has been published by JENSEN and MECKLING (1976). Losely following their approach we will analyse the conflict between managers, on the one hand, who hold certain stakes in a firm's equity and, on the other hand, capital owners who perfectly delegate decision making to managers; in a first attempt we assume that there is no debt incurred, yet. We consider the manager's wage income to reduce total payments to the owners as a whole, i.e. the total return to the owners in $\tau = 1$ is

$$z_1(a,.) - h_1(a,.) \ .$$

Thus, a manager's objective function may be stated as

$$h_o(a) + \pi(h_1(a,.)) + \alpha(z_o(a) - h_o(a) + p_E(a)) \ .$$

Supposing the market price of equity to reflect the manager's decision properly, implies:

$$p_E(a) = \pi(z_1(a,.) - h_1(a,.)) \ .$$

If, as JENSEN and MECKLING assume, the manager is the sole proprietor in the jump-off state his/her objective function is ($\alpha = 1!$):

$$h_o(a) + \pi(h_1(a,.)) + z_o(a) - h_o(a) + \pi(z_1(a,.) - h_1(a,.)) =$$

$$= z_o(a) + \pi(z_1(a,.)) = p_G(a)$$

i.e. s/he maximizes the total value of his/her firm.

If s/he is not the sole owner, initially, we have:

$$h_o(a) + \pi(h_1(a,.)) + \alpha\left[z_o(a) - h_o(a) + \pi(z_1(a,.)) - \pi(h_1(a,.))\right]$$

$$= \left[h_o(a) + \pi(h_1(a,.))\right](1-\alpha) + \alpha\left[z_o(a) + \pi(z_1(a,.))\right]$$

i.e. s/he maximizes a convex combination of wage income and of total (gross) value before deduction of wage income. This latter objective function reflects his/her interests immediately after having sold off a part of his/her stake being a hundred per cent owner up to that point in time. Comparing the two objective function shows that there is an incentive to change the initial policy of maximizing the firm's total market value after having sold off a part of the firm's equity.

Contracting for a wage scheme

$$h_o(a) = \gamma \cdot z_o(a) \text{ and } h_1(a,.) = \gamma \cdot z_1(a,.) \qquad (\text{for some } \gamma > 0)$$

seems to be a promising way to eliminate this incentive that, obviously, must result in a reduction of the initial market value if it is anticipated by external owners. The resulting conformity of interests is, however, unstable as the following consideration will show:

It should be accounted for the manager who is entitled with dispository power, to be able to enhance his/her income above the contractually agreed level by perk consumption (underhand withdrawals, shirking); in this case his/her total wage income consists of a contractually fixed part $(h_o^*(a), h_1^*(a,.))$ and an amount $(c_o^P(a), c_1^P(a,.))$ obtained by perk consumption. Disguised perk consumption will, however, only be successful if it is practised during the respective accounting period by reducing the volume of capital expenditures little by little, rather than by a withdrawal from gross revenues at $\tau = 1$. Thus it seems reasonable to take into account that the gross profit, by itself, will be affected by perk consumption:

$$z_1^*(a,.) = z_1(a, c^P, .) \ .$$

Now the manager's objective function reads:

$$\{[h_o^*(a) + \pi(h_1^*(a,.))] + [c_o^P(a) + \pi(c_1^P(a,.))]\}(1-\alpha)$$

$$+ \alpha[z_o(a) + \pi(z_1^*(a,c^P,.))] \longrightarrow \max!$$

Again, it is seen very clearly that a solely owning manager $(\alpha = 1!)$ will maximize the firm's total value; particularly, s/he won't give way to the illicit practice of perk consumption, provided his/her decisions are anticipated by the market early enough. But as soon as s/he *has* sold off a part of his/her business his/her concern will change: the relevance of firm's gross market value will be restricted to α per cent of it (from that on) whereas the sum of contractually agreed upon income and perk consumption will gain a $1 - \alpha$ per cent relevance. Contractually relating his/her wage income to the firm's gross profits, however, yields the following objective function:

$$[c_o^P(a) + \pi(c_1^P(a,.))](1-\alpha) + [\alpha + \gamma(1-\alpha)]\,[z_o(a) + \pi\,(z_1^*(a,c^P,.))] \longrightarrow \max\,!$$

In contrast, the objective function of the external owners may be written as

$$z_o(a) + \pi\,(z_1^*(a,c^P,.)) - [h_o^*(a) + \pi(h_1^*(a,.))] - [c_o^P(a) + \pi(c_1^P(a,.))] \longrightarrow \max\,!$$

or, adjusting for the compensation scheme:

$$(1-\gamma)[z_o(a) + \pi(z_1^*(a,c^P,.)] - [c_o^P(a) + \pi\,(c_1^P(a,.))] \longrightarrow \max\,!$$

The potential conflict of interest between the manager and external owners (having no dispository power) due to the possibility of perk consumption is evident: although c.p. both of them prefer a higher gross market value of the firm to a lower one a manager will not ignore that he is able to benefit from perk consumption at the cost of the external owners. They clearly prefer zero perk consumption whereas a manager's attitude is ambivalent in that he must settle a trade-off between income enhancement via perk consumption and a thereby induced income shortening via reduced market value; the manager's relative preference for perk consumption may be evaluated as:

$$\frac{1-\alpha}{1-\alpha+\alpha+\gamma(1-\alpha)} = \frac{1-\alpha}{1+\gamma(1-\alpha)} = \frac{1}{\gamma+(1-\alpha)^{-1}}$$

It will be stronger the smaller the manager's stake in equity will be.

A first step in limiting the conflict of interest between a manager and the external owners is the agreement on a success-oriented compensation scheme; any remaining disharmonies due to the possibility of perk consumption may be dealt with in terms of appropriate contractual arrangements (limitation of permissible dispositions) and adequate auditing. It should be recognized that the limitation of competences as well as auditing is costly, directly, or in that it may give rise to suboptimal decisions. These effects should be balanced against the advantages of delegating decision making. Furthermore, another related potential source of conflict deserves recognition: a manager's incentive for perk consumption becomes stronger the more he reduces his share in total equity, thus is may prove reasonable to stipulate by contract that he is bound to hold no less than a particular share in the equity of the firm. As a consequence of Assumption 5 which ensures the possibility to reproduce all the the cash-flows that may be generated by any combination of claims on the respective firm ('spanning'), such a stipulation does not cause welfare losses in our setting; the only effect is reducing the manager's possibilities to reach a position which gives rise to increasing incentives for perk consumption (FAMA (1980) focuses on disciplining effects of a market for management services with respect to manager's behaviour).

3.3 The Relation Between Managers, Owners, and Creditors

Finally we will conciliate the thus far separately considered cases of owner-creditor and owner-manager conflict.

Denote by

$z_1^*(a,c^P,.)$	the gross cash-flow after deduction of the manager's perk consumption
h_τ^*	the contractually agreed upon wage income of the manager ($\tau = 0,1$)
c_τ^P	the amount of the manager's perk consumption
$f(a,c^P,.)$	the cash-flow received by the creditors.

The owners receive the following payments

$$e_o(a) := z_o(a) - h_o^*(a) - c_o^p \qquad \text{in } \tau = 0$$

and

$$e_1(a,c^p,.) := z_1^*(a,c^p,.) - h_1^*(a,.) - c_1^p(a,.) - f(a,c^p,.) \qquad \text{in } \tau = 1$$

A manager who has not lent any money to his/her firm pursues the following objective function:

$$\{[h_o^*(a) + \pi(h_1^*(a,.))] + [c_o^p(a) + \pi(c_1^p(a,.))]\}(1-\alpha)$$

$$+ \alpha[z_o(a) + \pi(z_1^*(a,c^p,.)) - \pi(f(a,c^p,.))] \longrightarrow \text{max!}$$

where α is his/her initial share in the respective firm's equity.

The objective function of the external owners may be summarized as:

$$e_o(a) + \pi(e_1(a,c^p,.)) =$$

$$= [z_o(a) + \pi(z_1^*(a,c^p,.))] - [h_o^*(a) + \pi(h_1^*(a,.))]$$

$$- [c_o^p + \pi(c_1^p(a,.))] - \pi(f(a,c^p,.)) \longrightarrow \text{max!}$$

Besides the owners the creditors are also affected by the manager's incentive for perk consumption to the extent that the probability of default increases and the total value of assets in case of insolvency is reduced. In addition, the incentive to take actions designed to maximize the market value of equity at the cost of the creditors' claims - which emerges from the, at least partial, coincidence of management and ownership - will generally also do them harm. Thus, we have a conflict between insiders (managers) and outsiders (external owners and creditors) on one hand, and a conflict between owners and creditors on the other hand.

Reducing the potential for conflict seems possible by the following compensation scheme:

$$h_o^*(a) = \gamma z_o(a)$$

$$h_1^*(a,.) = \gamma z_1^*(a,c^p,.)$$

In consequence, the manager's objective function becomes:

$$\left[c_o^p(a) + \pi(c_1^p(a,.))\right](1-\alpha) + (\alpha+\gamma(1-\alpha))\left[z_o(a) + \pi(z_1^*(a,c^p,.))\right] -$$

$$- \alpha\pi(f(a,c^p,.)) \longrightarrow \max!$$

whereas the owners prefer according to:

$$(1-\gamma)\left[z_o(a) + \pi(z_1^*(a,c^p,.))\right] - \left[c_o^p + \pi(c_1^p(a,.))\right] - \pi(f(a,c^p,.)) \longrightarrow \max!$$

The manager's and the owners' concerns would be, now, in full harmony if not reducing his/her share in equity would produce stronger incentives for the manager to enhance his/her perk consumption and weaker incentives to increase the market value of equity at the cost of the creditors.

Taking into account that there may be reasons not to allow a manager to specialize to holding some share in the *equity* of his/her firm we may restate his objective function as follows

$$\left[c_o^p(a) + \pi(c_1^p(a,.))\right](1-\alpha) + (\alpha+\gamma(1-\alpha))\left[z_o(a) + \pi(z_1^*(a,c^p,.))\right] -$$

$$- (\alpha-\beta)\pi(f(a,c^p,.)) \longrightarrow \max!$$

where β indicates his/her share in the firm's total debt.

Now there are two strategies to eliminate the incentive to impair the creditors for the owners' sake:

- not to allow the manager to buy any of his firm's securities ($\alpha = \beta = 0$),
- prescribing proportional participation ($\alpha = \beta$) by contract.

Supposed that there are effective control mechanisms to ensure contract enforcement, in both cases the manager's objective function reduces to:

$$\left[c_o^p(a) + \pi(c_1^p(a,.))\right](1-\alpha) + (\alpha+\gamma(1-\alpha))\left[z_o(a) + \pi(z_1^*(a,c^p,.))\right] ,$$

which results in an actual elimination of the incentive to impair the creditors. The incentive for perk consumption is, however, very strong in case of $\alpha = 0$, thus imposing costs on all external parties concerned: therefore, restricting and controlling perk consumption requires special attention in this case. Appropriate measures have been discussed above in the section separately dealing with the manager-owner conflict. However, urging the manager to hold a share of the firm's total debt ($\alpha = \beta \neq 0$) reduces the effectiveness of the creditors' claim that their demands should be met prior to those of other parties concerned. In this case the manager acts as an equally treated creditor by him/herself. In case the (normal) creditors stipulate me-first rules (or other protection mechanisms) they are likely to destroy the incentive effect related to the manager being a creditor.

4. Concluding Remarks and Some Topics for Future Research

This paper provided an approach to quantify the concerns of the different parties having stakes in a business activity, in terms of the implied consumption opportunities; it outlined inherent potentials for conflict and presents possible strategies to resolve the complex problem. We showed that the market value of the initial endowment with financial assets and labour contracts may be regarded as an indicator that allows for preference independent ordering of the desirability of implied consumption opportunity sets under particular assumptions on how financial markets work and which allocative features they have. Thus the market value of the initial endowment can be used as objective function serving as the basis for a quantitative analysis of the wellknown conflicts between owners and creditors, and managers and owners, respectively. This analysis particularly revealed the importance of an appropriate designed wage scheme for harmonizing the interests of managers and owners. A fully satisfactory solution for all the parties concerned seems however difficult to design because of a manager's incentive to impair both owners and creditors.

However, on a second view it becomes evident that the assumption justifying the maximization of the market value of the initial endowment - in particular the 'spanning' assumptions - presume that all stakeholders know the entrepreneurial action to be taken, in advance. This ensures, on the one hand, that the market prices of the respective financial assets properly reflect the expected consequences and, on the other hand, that the hedging portfolios, needed for separation, can actually be established. This is not compatible with the realistic guess that there usually *is* earlier and more detailed information on corporate policy and its consequences for insiders. The fact that one should not expect to obtain separation results in such an economic environment since they require perfect information on corporate policy on the part of all stakeholders, causes serious problems in formulating the different involved parties' objective functions and thus casts considerable doubt on the applicability of approaches dealing with asymmetrical information by means of maximization of market values (cf.: among others MYERS/MAJLUF (1984)).

5. Appendices

5.1 Appendix I: On the Mathematical Structure of Consumption Opportunity Sets

The definition of consumption opportunity sets suggests constructing the following mapping θ_a:

$$\theta_a : \mathbb{R}^{m+3} \longrightarrow \mathbb{R} \times X_\Omega$$

with

$$\theta_a(\alpha^n, \beta^n, y^n, x^n) = \begin{pmatrix} -\alpha^n p_E - \beta^n p_D - P \cdot y^n - x^n \\ \\ \alpha^n \max\{z_1(a,.) - D, 0\} + \beta^n \min\{z_1(a,.), D\} + Z \cdot y^n + R \cdot x^n \end{pmatrix}$$

obviously θ_a is a linear mapping for each a the image of which, Im θ_a, is a linear subspace of $\mathbb{R} \times X_\Omega$. Now the consumption opportunity set may formally be defined as follows:

$$C(a;\alpha,\beta,y,x) = \left[\begin{pmatrix} h_o(a) + \alpha(z_o+p_E) + \beta p_D + P\bullet y + x \\ \\ h_1(a,.) \end{pmatrix} + \text{Im}\ \theta_a\right] \cap \mathbb{R}_+ \times X_\Omega$$

Thus, the term in square brackets represents a linear manifold consisting of the linear subspace Im θ_a displaced by the vector in the inner brackets. It should be recognized that the image Im θ_a does not depend on the initial endowment of the respective individual. The first term in square brackets indicates how the consumption possibilities depend on the initial endowment. $\mathbb{R}_+ \times X_\Omega$ represents a non-negativity constraint in period $\tau = 0$. The following simple consideration may be helpful. Look at Figure 1 in the body of the paper; the straight line in the (c_o, c_1)-diagram represents the image of θ_a shifted by the initial endowment as given by the vector in inner brackets; the broken line marks the section contradicting to the non-negativity constraint for c_o.

With $H(a) = \begin{pmatrix} h_o(a) \\ \\ h_1(a,.) \end{pmatrix}$ denoting the initial endowment

with respect to the wage income, and $I(a) = \begin{pmatrix} \alpha(z_o + p_E) + \beta p_D + P\bullet y + x \\ \\ 0 \end{pmatrix}$

denoting the initial endowment with financial assets, we have

$$C(a;\alpha,\beta,y,x) = \left[H(a) + I(a) + \text{Im}\ \theta_a\right] \cap \mathbb{R}_+ \times X_\Omega$$

This is a convex set. By a no-arbitrage condition (cf. WILHELM (1985)) this set must not intersect the following set

$$H(a) + I(a) + \mathbb{R} \times X_\Omega^+$$

except for the point $H(a) + I(a)$. This latter set is convex as well. Now there exists as separating hyperplane which separates $C(a;\alpha,\beta,y,x)$ from $H(a) + I(a) + \mathbb{R}_+ \times X_\Omega$. This hyperplane defines one possible version of the positive price functional π having the above-mentioned representation by a positive (stochastic) discount factor q.

5.2 Appendix II: Proof of Separation I

The consumption opportunity set may be written as

$$C(a;\alpha,\beta,y,x) = \left[H(a) + I(a) + \mathrm{Im}\ \theta_a\right] \cap \mathbb{R}_+ \times X_\Omega$$

For a proof of the first part of the separation result it is sufficient to show that $\mathrm{Im}\ \theta_a$ does not depend on corporate policy. Let be $\binom{v}{w} \in \mathrm{Im}\ \theta_a$, i.e.

$$v = -\alpha^n p_E - \beta^n p_D - P \cdot y^n - x^n$$

$$w = \alpha^n \max\{z_1(a,.) - D, 0\} + \beta^n \min\{z_1(a,.), D\} + Zy^n + R \cdot x^n$$

By Assumption 5 we conclude:

$$w = \alpha^n \cdot Z \cdot y_E^a + \beta^n \cdot Z \cdot y_D^a + Zy^n + R \cdot x^n$$

$$= Z \cdot \{\alpha^n y_E^a + \beta^n y_D^a + y^n\} + R \cdot x^n$$

Using Assumption 2 and 3 we get:

$$v = -\left[P \cdot \{\alpha^n y_E^a + \beta^n y_D^a + y^n\} + x^n\right],$$

where Z,R and P do not depend on a. We construct the mapping

$$\theta : \mathbb{R}^{m+1} \longrightarrow \mathbb{R} \times X_\Omega$$

by

$$\theta(y^*, x^*) = \begin{pmatrix} -Py^* - x^* \\ \\ Zy^* + R\,x^* \end{pmatrix}, \quad y^* \in \mathbb{R}^m,\ x^* \in \mathbb{R}$$

From what we have just derived we immediately see $\mathrm{Im}\ \theta_a = \mathrm{Im}\ \theta$.

The second part of assertion may be concluded as follows: assume $H(a) = H(b)$ but $I(A) < I(b)$ to hold and let further be

$$\begin{pmatrix} c_o(a) \\ \\ c_1(a,.) \end{pmatrix} = H(a) + I(a) + \theta(y_a, x_a) \in \mathbb{R}_+ \times X_\Omega$$

We choose

$$\begin{pmatrix} c_o(b) \\ \\ c_1(b,.) \end{pmatrix} = H(b) + I(b) + \theta(y_a,x_a) > H(a) + I(b) + \theta(y_a,x_a) = \begin{pmatrix} c_o(a) \\ \\ c_1(a,.) \end{pmatrix}$$

therefore getting

$$\begin{pmatrix} c_o(b) \\ \\ c_1(b,.) \end{pmatrix} > \begin{pmatrix} c_o(a) \\ \\ c_1(a,.) \end{pmatrix} \quad ;$$

which implies that the condition $c_o(b) \geq 0$ is met, thus the consumption opportunity set attached to b dominates that one attached to a.

5.3 Appendix III: Proof of Separation II

The following holds true with respect to any consumption opportunity set.

$$C(a;\alpha,\beta,y,x) \quad = \quad \left[H(a) + I(a) + \mathrm{Im}\,\theta\right] \cap \mathbb{R}_+ \times X_\Omega$$

We consider

$$H(a) + \mathrm{Im}\,\theta = \left\{ \begin{pmatrix} h_o(a) \\ \\ h_1(a,.) \end{pmatrix} + \theta(y^*,x^*) \mid y^* \in \mathbb{R}^m, x^* \in \mathbb{R} \right\}$$

$$= \left\{ \begin{pmatrix} h_o(a) \\ \\ Z \cdot y_A^a \end{pmatrix} + \theta(y^*,x^*) \mid y^* \in \mathbb{R}^m, x^* \in \mathbb{R} \right\}$$

$$= \left\{ \begin{pmatrix} h_o(a) \\ \\ Z \cdot y_A^a \end{pmatrix} + \begin{pmatrix} -Py^* - x^* \\ \\ Zy^* + R \cdot x^* \end{pmatrix} \;\middle|\; y^* \in \mathbb{R}^m, x^* \in \mathbb{R} \right\}$$

$$= \left\{ \begin{pmatrix} h_o(a) + \pi(h_1(a,.)) \\ \\ 0 \end{pmatrix} + \begin{pmatrix} -Py^* - Py_A^a - x^* \\ \\ Z \cdot y_A^a + Zy^* + R \cdot x^* \end{pmatrix} \;\middle|\; y^* \in \mathbb{R}^m, x^* \in \mathbb{R} \right\}$$

$$= \begin{pmatrix} h_o(a) + \pi(h_1(a,.)) \\ 0 \end{pmatrix} + \left\{ \begin{pmatrix} -P(y^* + y_A^a) - x^* \\ Z(y^* + y_A^a) + R \cdot x^* \end{pmatrix} \middle| \; y^* \in \mathbb{R}^m, x^* \in \mathbb{R} \right\}$$

Of course we have

$$\left\{ \begin{pmatrix} -P(y^* + y_A^a) - x^* \\ Z(y^* + y_A^a) + R \cdot x^* \end{pmatrix} \middle| \; y^* \in \mathbb{R}^m, x^* \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} -Py^{**} - x^* \\ Zy^{**} + R \cdot x^* \end{pmatrix} \middle| \; y^{**} \in \mathbb{R}^m, x^* \in \mathbb{R} \right\} = \mathrm{Im}\,\theta$$

Therefore $H(a) + \mathrm{Im}\,\theta = \pi(H(a)) + \mathrm{Im}\,\theta$ holds meaning that the consumption opportunity set depends on corporate policy only via

$$h_o(a) + \pi(h_1(a,.)) + \alpha(z_o(a) + p_E) + \beta p_D + P \cdot y + x$$

Thus we have proved the first part of Separation II, the proof of the second part proceeds in the same way as for Separation I.

References *)

Brito, N.O. (1977): Marketability restrictions and the valuation of capital assets under uncertainty, JoF 27, 1109 f

DeAngelo, H. (1981): Competition and unanimity, AER 71, 18 f

Fama, E.F. (1980): Agency problems and the theory of the firm, JPE 88, 288 f

Harrison, J.M.; Kreps, D.M. (1979): Martingales and arbitrage in multiperiod securities markets, JET 20, 381 f

Haugen, R.A.; Senbet, L.W. (1981): Resolving the agency problems of external capital through options, JoF 36, 629 f

Jensen, M.C.; Meckling, W.H. (1976): Theory of the firm: managerial behaviour, agency costs and ownership structure, JFE 3, 305 f

Kudla, R.J. (1984): A note on resolving agency problems through warrants, JBF&A 11, 67 f

Makowski, L.; Pepall, L.(1985): Easy proofs of unanimity and optimality without spanning: a pedagogical note, JoF 40, 1245 f

Mayers, D. (1972): Nonmarketable assets and capital market equilibrium under uncertainty, in: M.C. Jensen (ed), Studies in the Theory of Capital Markets, New York - Washington - London, 223-248

Myers, St.C.; Majluf, N.S. (1984): Corporate financing and investment decisions when firms have information that investors do not have, JFE 13, 187 f

Ross, S.A. (1978): A simple approach to the valuation of risky streams, JoB 51, 453 f

Smith, C.W.; Warner, J.B. (1979): On financial contracting an analysis of bond covenants, JFE 7, 115 f

Wilhelm, J. (1983): Marktwertmaximierung - Ein didaktisch einfacher Zugang zu einem Grundlagenproblem der Investitions- und Finanzierungstheorie, ZfB 53, 516 f

Wilhelm, J. (1985): Arbitrage Theory, Berlin et al.

*) Abbreviations: JoF Journal of Finance, AER American Economic Review, JPE Journal of Political Economy, JET Journal of Economic Theory, JFE Journal of Financial Economics, JBF&A Journal of Business Finance and Accounting, JoB Journal of Business, ZfB Zeitschrift für Betriebswirtschaft.

Section 4 Financial Contracting and Dividends

Signalling and Market Behavior

Andreas Gruber

Summary: This paper represents and examines a model which shows possible effects of heterogeneous expectations on the price level of a share on a stock exchange. It will be shown that market prices free of arbitrage can be higher than the present values of expected future dividend payments of all investors. This result is based on the assumption that some investors signal their expectations to inform other investors. Motivational reasons for these signals are investigated: There are cases with strong arguments for signalling in theory and practice. Finally it is shown that investors are generally motivated to advice other investors only if either their own transactions have already been executed or these advices are not free of charge.

Table of Contents:

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

1. Introduction

There are two procedures for predicting future stock prices in practice: The fundamental and the technical analysis. This paper reviews a generalization of both strategies. Assuming heterogeneous expectations Harrison and Kreps (1978) developed a formal model which includes expectations of other investors in the investment strategies. Although this HK-model is not very useful for real investments because there will be problems in getting the necessary information, it gives some interesting insights into the psychology of the market.

The basic idea is not new. Keynes indicated with his view of a beauty contest, that assets won't be valued only by known or expected fundamental factors, but also indirectly by the expectation "what average opinion expects average opinion to be" (Keynes (1936), p. 156).

The paper has four parts. Section 2 is a review of the Harrison-Kreps-model. In the HK-model every investor estimates the value of a share as discounted rate of all expected payments including dividends and the selling price. Prices are endogenous, and it is assumed initially that investors send signals about their own subjective expectations on future dividends of the inspected stock. With these signals, investors are able to consider opportunities to sale in their calculations. Orosel (1985a) calls this the keynesian approach. The antithesis to this is the neoclassical or standard approach: Stock prices are determined by the expected stream of future dividends only. If expectations are homogeneous, these two views are identical.

Section 3 examines the main assumption of the HK-model: All investors signal their expectations on future dividends to inform other investors, so that they are able to use them in their investment strategies. A surprising result is shown: This assumption is not restrictive. Investors really are motivated to signal their expectations.

In section 4 it is shown that this motivation does not hold if other investors are able to take these expectations over and copy the signalled strategy. In general investors are only motivated to advice other investors, if these recommendations are not free of charge or if the adviser has already executed the recommended transactions for himself.

2. A Review of the Harrison-Kreps-Model

2.1 Assumptions

The following assumptions are necessary for the model which shows the effects of the keynesian strategy on the market price. Some of these assumptions will be modified later.

(A1) The market for a single stock will be investigated. Trading takes place only at a discrete sequence of time $t \in \mathbb{N}_o$.

(A2) Each stock yields dividends $d_t \geqq 0$ in every period t (=[t-1,t)). The dividend for period t will be announced and paid right before trading at t.

(A3) Ex ante dividends are understood as a random process

$$d_t : \Omega \to \mathbb{N}_o , \quad w \to d_t(w) .$$

Ω is the set of all relevant events w, which could influence the dividend payments, for example earnings, productivity, plans in the future, economic state, rate of exchange etc.

(A4) Every investor has his own subjective distribution function for the exogenously determined random process $\{d_t\}$. Investors with the same distribution function are summarized in classes. There are only two distinct investor classes I and J.

Assumption (A4) states that trading investors have no influence on the performance of the corporation or the dividend payments. In some corporations there are a few stockholders having a majority in the annual general meeting. They do not influence the market price of the shares directly because they keep their investments constant. Thus the market price is determined by the trading of investors who have no influence on the performance of the corporation.

The expectations of investors on future dividend payments depend on information about the performance of the corporation. We assume for simplification:

(A5) The only information investors use to calculate the expected dividend payment of period t is the dividend payment of the preceding period t-1.

All relevant information for the valuation of the performance of the corporation is reflected by the dividend payment in the corresponding period. Harrison and Kreps (1978) argue that there is no difficulty in defining information as a random vector, the realizations of which determine the dividends in a definite matter. It is important that all these data become public only after investors are trading in the corresponding period: Investors don't have any information on the performance of the corporation in the current period, but the have all relevant information (fundamental facts, past stock prices) on it in preceding periods. This homogeneous information will lead to heterogeneous expectations by evaluating them differently.

(A6) All investors are risk-neutral. They are maximizing the present value of their expected payments from the shares. Future payments are discounted by q^{-1} per period.

(A7) The organisation of the market is: All investors come together at one place, the stock exchange. An auctioneer (the stockbroker) collects all orders publicly and, when no more orders are given, sets up a price, so that the turnover of shares is the highest possible. Investors are allowed to cancel their orders before rate fixing.

(A8) There are no transaction costs.

(A9) The supply of shares is assumed to be constant.

(A10) Short sales are not permitted.

(A11) Every investor signals his own subjective distribution function about future dividends. These signals are free of charge.

Therefore every investor knows the expectations and strategies of all other investors.

(A12) All these probability distributions are time independent. There are no learning skills.

(A13) There are no monetary restrictions on the investor classes.

(A14) There is competition between the different investor classes and within each class.

With these assumptions it is easy to give a first statement about the market value of a share.

Lemma 1: In each state d_t this class of investors is buying all the shares which assigns the highest value to them. The market price p_t is equal to the present value of the future returns of a share expected by this class of investors.

The reason is: No investor can buy shares for a lower price because the market is organized similar to an auction (A7). The market price will be forced up because of internal competition in each class (A14) and unrestricted budgets (A13).

In contradiction to some of these assumptions in reality there are many more than two investor classes in a stock market. In addition budgets are restricted. Including these more general assumptions in the model the qualitative results would stay the same without reservation. But the theory would be much more complicated, because it would be necessary to calculate the market price on the basis of an additional equilibrium concept, which depends on the expectations of all investors and the corresponding budgets.

2.2 The Price Process

Lemma 1 is the starting point for further analysis. First of all it is necessary to specify this statement on the market price. Since the decisions of investors are based on rationality (A6), the only question is: How can investors estimate the present value of the payments from a share. On the basis of their own distribution function it is easy for them to compute discounted values of expected dividends. Additionally they have to consider a selling strategy, which includes the estimation of the selling-price and the timing of the sale. Selling strategies depend on buying strategies of other investors. These depend on their expectation about the trend of prices, i.e. on their own selling strategies.

The analysis to solve this dilemma is split up in two parts. First a desirable price scheme, a sequence of prices $\{p_t\}$, will be stated, dicussed and then simplified. Then it will be shown that it is for investors possible to compute these prices ex ante : Investors have all the information they need for their investments. In this context a price scheme, which is free of arbitrage (over time) for all investors, is desirable. This condition guarantees that no investor is motivated to cause transactions which would change at least one price in the price scheme.

<u>Lemma 2:</u> A price scheme $\{p_t\}$ is free of arbitrage for all investors, if and only if

$$p_t(d_t) = \max_{S=I,J} \sup_{T} E^S \left(\sum_{k=t+1}^{T} q^{-(k-t)} \cdot \tilde{d}_k + q^{-(T-t)} \cdot p_T(\tilde{d}_T) \mid \tilde{d}_t = d_t \right)$$

for all t, $d_t \in \mathbb{N}_o$. (1)

The reason is: Both investor classes value a share as the sum of dicounted expected dividend disbursements before selling the share in T plus the discounted expected selling-price. For this every investor has to estimate his optimal selling time T. Using Lemma 1 the market price is the maximum of these present values. If the market price would be less (or greater) at least one investor could get positive returns by arbitrage.

Harrison/Kreps (1978) simplified condition (1):

<u>Theorem 1:</u> Condition (1) is equivalent to

$$p_t(d_t) = \max_{S=I,J} q^{-1} \cdot E^S\left(\tilde{d}_{t+1} + p_{t+1}(\tilde{d}_{t+1}) \mid \tilde{d}_t = d_t\right)$$

for all t, $d_t \in N_o$. (2)

The reason is: Investing in condition (1) fulfilling shares is a fair game for all $t \in \mathbb{N}_o$. The price reflects always all available information. Therefore in (2) the estimated market price p_{t+1} includes all expectations about future dividends and market prices. A further investigation of future prices can't improve the investment decision.

The formal proof introduces the prices p_t as submartingales; T is a stopping time. Then Doob's optional sampling theorem is used (Harrison/Kreps (1978), p. 330). Another way to proof this theorem could be based on an idea of Samuelson (1965). Using a similar approach he showed that properly anticipated prices fluctuate randomly. The future price sequence is, subject to the axiom of rational expectations, a fair game.

The equilibrium concept in Lemma 2 and Theorem 1 can be used as a mark for the quality of a given price scheme. But it doesn't give us the possibility to compute a price scheme on the basis of the distribution functions. An algorithm to do this is also given by Harrison/Kreps (1978):

Define $p_t^o(d_t) = 0$ and

$$p_t^n(d_t) = \max_{S=I,J} q^{-1} \cdot E^S\left(\tilde{d}_{t+1} + p_{t+1}^{n-1}(\tilde{d}_{t+1}) \mid \tilde{d}_t = d_t\right) \qquad (3)$$

for all $n \in \mathbb{N}$, t, $d_t \in \mathbb{N}_o$.

Because $p_t^n(d_t)$ is a non-negative, monotonously increasing sequence, we conclude

$$p_t^*:= \lim_{n\to\infty} p_t^n(d_t) = \begin{cases} c \geqq 0 \\ \infty \end{cases} . \qquad (4)$$

Now it is possible to show:

<u>Theorem 2:</u> a) $\{p_t^*\}$ is free of arbitrage.

b) If $\{p_t\}$ is free of arbitrage, then

$$p_t^*(d_t) \leqq p_t(d_t) \qquad \text{for all } t,\ d_t \in \mathbb{N}_o.$$

Again this Theorem has it's origin in Harrison/Kreps (1978). The proof is also based on Doob's optional sampling theorem.

Theorem 2 asserts not only that the a priori construction of a price scheme free of arbitrage is possible, but also that the so constructed one is minimal. This suits the careful practices of investors on the stock market. However, in (A6) investors are assumed to be risk-neutral. The reason for this is to keep the analysis easier to survey. The qualitative results would be the same with risk-averse investors.

Non-minimal price schemes free of arbitrage are known. Remember the saying: 'The hausse bites the hausse'. In a hausse stock prices are rising not for fundamental reasons. The demand is immense only because investors expect stock prices to increase further.

We conclude that the condition of a price scheme free of arbitrage is weaker than the condition of rational expectations. If investors are acting rationally by maximizing their expected payments, the price scheme will be free of arbitrage; otherwise at least one investor could yield profits by arbitrage. But the behavior of rational expectations is not required for a price scheme free of arbitrage. For example the price scheme $\{p_t\}$ with $p_t = \infty$ for all $t \in \mathbb{N}_0$ is free of arbitrage, but it doesn't reflect always rational expectations. Investors proceed with rational expectations in the sense of Grossman (1981), if a minimal price scheme free of arbitrage is given. This is not valid, when there are no selling possibilities known, for example because investors don't signal their expectations or their subjective distribution function respectively. In this case prices are fixed only by the investors own expectations on future dividends. This price fixing process meets the conditions of a Walrasian equilibrium (Grossman (1981)), because investors don't try to get information about the expectations of other investors from the market price.

Prices free of arbitrage have two interesting properties:

<u>Corollary 1:</u> If $\{p_t\}$ is a price scheme free of arbitrage, then

$$p_t(d_t) \geqq E^S\left(\sum_{k=t+1}^{\infty} q^{-(k-t)} \cdot \tilde{d}_k \mid \tilde{d}_t = d_t\right) \qquad (5)$$

for all t, $d_t \in \mathbb{N}_0$, $S=I,J.$.

This statement follows from Theorem 1. It says, that prices free of arbitrage are greater than or at least equal to the present values of the expected dividends of all investors.

<u>Corollary 2:</u> If investors have homogeneous expectations, then

$$p^*_t(d_t) = E^S\left(\sum_{k=t+1}^{\infty} q^{-(k-t)} \cdot \tilde{d}_t \mid \tilde{d}_t = d_t \right) \qquad (6)$$

for all t, $d_t \in N_o$, S=I,J.

This Corollary follows immediately from equation (3) and Theorem 2. Using the minimal price scheme it shows that homogeneous expectations are special case of heterogeneous expectations. Or more generally, the neoclassical view of Orosel (1985a) is only a special case of his so-called keynesian view (see Section 1. Introduction). The more heterogeneous expectations of investors are, the less depend their strategies on their own expectations on future dividends; and the more they depend on expected variations in the market price, i.e. on the expectations of all other investors.

3. Extension and Signals

It was assumed in (A11) that all investors signal their expectations even before they use them for their own trading activities. It is not obvious why they should be motivated to do this. Therefore first some qualitative arguments for signalling will be listed. Then we show in the model that investors really can get non-negative returns by signalling their expectations. Spence (1974) used the term 'signalling' in a slightly different context. He argued that signalling gives the possibility to split an otherwise pooled market. In the HK-model, signals will not split a pooled market. Rather than that, they help to raise returns in the descibed two-person zero sum game.

3.1 General Incentives for Signalling

The following reasons for signalling the own investment strategy are conceivable:

(I1) Gain of prestige because of accurate forecasts:
Brokers publicate sample deposits in the conviction, that these recommendations have outstanding returns.

The prestige of the broker can grow with these signals and so new customers can be found.

(I2) Principal agent relations:
Investors like to control the performance of the managers of their property. They advise the managers to signal their efforts or achieved results.

(I3) Exchange of experience:
Investors join investment clubs in the hope to get better returns from their investments when they discuss them together.

(I4) Lower transaction costs:
At small stock exchanges like Stuttgart/FRG professional investors give their orders publicly by shouts, because this is easier and faster.

(I5) Self fulfilling prophecies:
Sometimes an investor can manage by signalling, that other investors copy his strategy or at least adjust their own strategies to this one. Usually these signals are sent when the corresponding transaction already are executed. If other investors then order similar transactions, this lowers the risk of the first investment and helps to fulfill the prophecy. But the later the respective investment is done, the higher is the risk in it.

(I6) Profits by market reactions:
With additional information other investors will be able to reconsider their investment strategies. This might have reactions on market prices. Possibly the signalling investor can foresee these reactions and consider them in his investment decisions. This is similar with inventives (Hirshleifer (1971)).

Both incentives (I5) and (I6) will be examined further.

3.2 Signals and Expected Returns

This section examines whether it can be possible for investors to get positive returns from market reactions that are induced by their own signals. Again, suppose that no investor will change his subjective distribution function on future dividend payments because of signals. Investors use them only for computing opportunities to sell shares. First a special case is represented in detail, subsequently results of other cases are listed. We assume instead of (A11):

(A11') Investors I and J do not signal their expectations on future dividends.

The point of the question is: Under which circumstances are investors of class I motivated to signal their expectations? To solve this, first we develop price schemes free of arbitrage for the two relevant cases, i.e. when nobody and when class I signals respectively. Second, with these two price schemes we're able to examine under which conditions class I is motivated to signal because of higher expected returns.

Lemma 3: If no investor signals his expectation on future dividend payments the price scheme $\{p_t\}$ is free of arbitrage, if

$$p_t(d_t) = \max_{S=I,J} E^S\left(\sum_{k=t+1}^{\infty} q^{-(k-t)} \cdot \tilde{d}_k \mid \tilde{d}_t = d_t\right) \qquad (7)$$

for all t, $d_t \in \mathbb{N}_0$.

Proof: The investment decisions of class I and J depend on their own expectations regarding future dividend payments only. Therefore the value of a share is in t

$$\text{for class I} \quad E^I\left(\sum_{k=t+1}^{\infty} q^{-(k-t)} \cdot \tilde{d}_k \mid \tilde{d}_t = d_t\right) =: E^I(d) , \qquad (8)$$

for class J accordingly.

The statement (7) follows from Lemma 1.

Lemma 4: Assume that investors of class I only signal their expectations. Then the price scheme $\{p_t\}$ is free of arbitrage, if condition

$$p_t(d_t) = \max \left\{ E^I\left(\sum_{k=t+1}^{\infty} q^{-(k-t)} \cdot \tilde{d}_k \mid \tilde{d}_t = d_t \right) \right. ,$$

$$\left. q^{-1} \cdot E^J\left(\tilde{d}_{t+1} + p_{t+1}(\tilde{d}_{t+1}) \mid \tilde{d}_t = d_t \right) \right\} \tag{9}$$

holds for all t, $d_t \in \mathbb{N}_o$.

Proof: The investors of class J are able to compute expected future market prices. Hence they can include selling possibilities in their strategies. For that reason the value of a share changes for the investors of J to

$$q^{-1} \cdot E^J\left(\tilde{d}_{t+1} + p_{t+1}(\tilde{d}_{t+1}) \mid \tilde{d}_t = d_t \right) =: E^J(d+p) . \tag{10}$$

The statement follows with Lemma 1.

The investors of class I will be motivated to signal their expectations in the state d_t, if their expected gains from signalling are positive or at least non-negative. Therefore their expected yields with and without signals have to be compared respectively. However, if class I would be motivated to signal, this is not equivalent with the fact that they really would get higher returns by signalling. On the one hand the realizations of dividends are a random process, on the other hand their subjective distribution function could be inappropriate to describe this process.

Suppose we are in state d_t, trading is not opened yet. There are two cases:

Case 1: $E^I(d) \leqq E^J(d)$

If class I does not signal the investors of class J buy all shares for the price $E^J(d)$ (This shall be also true, when the expected values are equal). If class I signals, the investors of J have to offer

$$E^J(d+p) \geqq E^J(d) \tag{11}$$

per share because of the knowledge of a favourable selling possibility in the future and because of their internal competition. If inequality (11) is strict and the investors of class I own at least one share, they are motivated to signal because of the higher selling-price. Without signals, the investors of class J get this surplus because the future selling possibility for J is given independently from the signals. If the expectations in (11) are equal, the signals have no influence on the market.

Case 2: $E^I(d) > E^J(d)$

If class I doesn't signal, they buy all shares for $E^I(d)$ with expected profit zero. This changes with signals from the investor class I only, if

$$E^J(d+p) > E^I(d) \quad . \tag{12}$$

In this case the investors of I are motivated to send signals because of the opportunity to sell shares with profit

$$E^J(d+p) - E^I(d) \tag{13}$$

per share.

Thus we have the following theorem.

Theorem 3: Assume that no investor signals his subjective distribution function on future dividend payments in the HK-model. Assume further that investors use signals to calculate future market prices only, but not to think over or revise their own subjective distribution function.
Then every investor is motivated to signal because of either higher or at least equal expected returns.

3.3 Further Results

Analogous it is possible to show the following results on motivations for signalling:

starting point	Interior competition and no budget restrictions in I and J	Interior competition only in J and/or budget restrictions only in I
Nobody signals	I or J are motivated respectively	I is motivated only in some cases, J always
I signals	J is motivated	J is motivated
J signals	I is motivated	I is motivated only in some cases

Table 1: Further results on motivations for signalling

This examination, however, could be based on a reverse formulation of the problem: Under which circumstances is an investor class motivated to search for information about expectations of other investors? Note that this is not equivalent to the question in which cases a sole investor is motivated to search for private information. Here it was indirectly assumed that information immediately becomes known publicly.

A short example will illustrate this: Suppose only one investor of a share demanding class has the insider information that there will be an excellent selling possibility in the future. Then he can buy as many shares as he wishes (assuming his budget is not restricted) just by offering a marginal higher price per share than the non-informed investors of his class. If he had not the information, he would have the same excellent selling possibility in the future. But he would not be able to buy the same amount of shares, because he would refuse to offer this higher price and therefore he would be in competition with other investors of his class. Not his profits per share, but his total profits, can grow with insider information in such a situation.

Contrary to this it can be shown that it is not profitable for the whole investor class to search for information on the expectations of other investors. Since this result seems to be surprising, the reason should be made more clear: It any information becomes known to many investors the prices will reflect them immediately. This is so, because not only prices, but already orders reflect information. At least in the model, but also in reality non-informed investors have the chance to adjust their strategies to these orders. Therefore investors have to pay their entire expected value for a share, when they are in competition with other homogeneous investors.

These results get some support from an empirical study of Mühlbradt (1978). He found that the New York Stock Exchange (NYSE) is not efficient in the strong form, but it is in the semi-strong form. This means that the use of public information for investments on the NYSE doesn't help to beat the market, i.e. to get mean returns higher than the market average.

3.4 Example

This example illustrates Theorem 3. Suppose there is a corporation ABC, whose dividend payments always are either \$ 0 or \$ 10. Two investor classes I and J are interested in buying shares of ABC. The investors of I believe that the corporation follows a policy of constant dividends. Their probability assessments are:

$$P^I(d_{t+1}=10|d_t=0) =: P^I(10|\ 0) = 0.1\ ; \quad P^I(0|\ 0) = 0.9\ ;$$

$$P^I(10|10) = 0.9\ ; \quad P^I(0|10) = 0.1\ .$$

The investor class J believes, that the dividend policy of ABC is more flexible:

$$P^J(10|\ 0) = 0.5\ ; \quad P^J(0|\ 0) = 0.5\ ;$$

$$P^J(10|10) = 0.6\ ; \quad P^J(0|10) = 0.4\ .$$

The discount-rate per period is assumed to be $q^{-1} = 0.9$ to reflect inflational tendencies. Lower rates could be used to reflect additionally risk-aversity regarding future pay-offs.

So far, no investor is signalling his expectations on dividend payments. They are operating only on the basis of their own expectations on dividend payments. In this situation investors of I reconsider, whether it wouldn't be better for them to signal their expectations.

They charge somebody who knows the expectation of the other investor class to investigate, whether they should signal or not. This person compares the expected returns of class I from investments with signalling with those without signalling.

Without signals the value of a share can be calculated for I and J in the following way:

$$E^I\left(\sum_{k=t+1}^{\infty} q^{-(k-t)} \cdot \tilde{d}_k \mid \tilde{d}_t=0\right) =:$$

$$E^I(d|\ 0) = 0.9 \cdot [0.9 \cdot E^I(d|0) + 0.1 \cdot 10 + 0.1 \cdot E^I(d|10)] \approx 32.14\ ,$$

$$E^I(d|10) = 0.9 \cdot [0.1 \cdot E^I(d|0) + 0.9 \cdot 10 + 0.9 \cdot E^I(d|10)] \approx 57.86\ ,$$

analogously follows $E^J(d|\ 0) \approx 49.45$,

$$E^J(d|10) \approx 50.44\ .$$

The investors of J rate the shares higher than the investors of I in the state with no dividend payments. This is reasonable, because they expect the resumption of dividend payments with higher probability. Therefore in this state the investor class J will buy all shares, in the other state the investors of I will repurchase them.

J's expected surplus per share in	$d_t=0$	is	49.45 - 49.45 = 0.00 ;
I's profit from selling a share in	$d_t=0$	is	49.45 - 32.14 = 17.31 ;
J's profit from selling a share in	$d_t=10$	is	57.86 - 50.44 = 7.42 ;
I's expected surplus per share in	$d_t=10$	is	57.86 - 57.86 = 0.00 .

The investors of class J get an unexpected surplus of 7.42 per share when they sell in the state of dividend payments. Again the reason is the internal competition in class I. If these investors knew of this favourable selling possibility before buying shares in $d_t=0$, they would have to consider it in their buying orders. Since there is internal competition in J the market price would rise, and this would be for the benefit of class I.

It seems that I is able to rise it's returns by signalling at the cost of J. To prove this suppose that I signals his intention to buy all shares back in the next state of dividend payments for a price of 57.86. With this, the value of a share changes for J in $d_t=0$ to:

$E^J(d|0) = 0.9 \cdot [0.5 \cdot E^J(d|0) + 0.5 \cdot 10 + 0.5 \cdot 57.86] \approx 55.52$.

If I is signalling, then

I's profit from selling a share in $d_t=0$ is 55.52 - 32.14 = 23.38 ;

J's profit from selling a share in $d_t=10$ is 0.00 .

The investors of class I are able to get the whole former surplus from J by signalling. This surplus, which was 7.42, has to be discounted and the random character of the payment has to be considered. Therefore I's surplus by signalling is only 6.07 per share.

The profound reason for this profit by signalling is due to heterogeneous expectations even after signalling. In $d_t=0$ investors of class I have pessimistic expectations on the near future. The only honest possibility for them to raise the expectations of other investors, and therefore their own earnings by selling shares, is to guarantee them to buy the shares back for a good price in the next period with dividend payments.

4. Switching Behavior

4.1 Example

The assumption (A7) has to be extended for this example, which is somehow contrary to the example in 3.4.

(A7') Additional to (A7) it is valid, that a sequence of prices is possible in one trading period t.

Again, there are two different investor classes I and J. In J internal competition is found, I exists only out of one investor i. If investor i values the shares higher than the investors in J he buys all shares. The price is less than his expected value of a share because he is in no competition with homogeneous investors. But as already mentioned investor i has to bear two kinds of risks (and chances): First the basis of his decision, his subjective distribution function, could describe the dividend process badly. And second, even when his distribution function is quite accurate, an unlikely sequence of dividend payments could harm his investments.

If investor i signals his expectations and if there are many other investors, who take them over and use them for investements by themselves, then investor i is able to tansfer his risky assets into secure ones by selling his shares. The price would equal his expected value, if these investors are also risk-neutral.

Since investor i is assumed to be risk-neutral, the reason for his profits is not his signalling, but the opportunity to buy all shares for an advantageous prices. The sale of the assets doesn't increase his utility. In this context only one possibility to increase investor i's utility by signalling is thinkable: He has to convince investors of his distribution function, who have more willingness to take over risks. In this case these investors will value the assets higher than investor i, although they are using the same subjective distribution function.

4.2 Incentives for Recommendations

On a stock market learning skills are very difficult to achieve, because you never know, whether your subjective expectations have been incorrect, or whether you just had poor luck, since an unlikely sequence of dividend payments has occured. Shefrin/Statman (1985) investigated this conflict first in a formal model and compared their results with empirical tests. Using the Prospect Theory (see Kahneman / Tversky (1979)) they showed, that investors usually realize positive returns on investments too early, because "they want to hasten the feeling of pride at having chosen correctly in the past" (p. 782). Contrary to this they ride loosers too long, because they are averse to loss realizations.

If the true present value of the dividend payments lies always between the investors expected present values, then investors will be disappointed on an average. This follows from Lemma 1: In every state all shares will be purchased by the most optimistic investor class for a price, which is higher than the value of the share. This is one of the reasons why investors might change their expectations or retire from investments in stocks.

There are other reasons for switching behavior like recommendations from a broker or advices from other competent persons. Using the model we show, under which circumstances investors could be motivated to advice other investors with the intention that they copy their strategies. To simplify matters we assume, that these advices (or signals) are free of charge.

The results and most important arguments are listed in table 2. It is assumed that some investors are changing from class I to class J.

Consequences on the structure of the investor classes	Implications on prices and expected profits of the former investors of J
No change of the structure (budget, internal competition) of both classes	No implications
J's expectations are greater than I's in t; less budget restrictions or stronger internal competition in J in t; no implications in the future	The market price increases in t => no benefits for J
J's expectations are less than I's in t; stronger budget restrictions or less internal competition in I in t; no implications in the future	The market price decreases in t => no benefits for J
J's expectations are greater than I's in t, but less in t+1; stronger budget restrictions or less internal competition in I in t+1	The market price decreases in t+1 and therefore also in t => no benefits for J in t+1 and in t (because of the worse selling possibility in t+1)
J's expectations are less than I's in t, but greater in t+1; less budget restrictions or stronger internal competition in J in t+1	The market price increases in t+1 and therefore also in t => no benefits for J if they own no shares before trading in t occurs; but if they do so, they get benefits, when their gains from the price increase in t are greater than their disadvantages from the price decrease in t+1.

Table 2: Consequences of recommendations

These are plausible results, because they support the following thesis: If someone is convinced of his strategy, he should not try to persuade other investors to order the same transactions before he has executed them for himself. Otherwise the whole cake must be divided between many hungry investors. This can be different with advices which are not free.

The existence of financial newspapers, which inform about performances of firms and give advices on buying and selling stocks, need not stand in contradiction with these results. An example, which was reported in the 'New York Times' of May 18, 1984, illustrates this (Orosel (1985a)): The author of the column 'Heard on the Street' in the 'Wall Street Journal' has an influence on the expectations of investors by publishing his investigations on the future performance of corporations. Since many readers believe these signals they change their expectations to be congruent with the authors ones. Knowing this, the author made a profit of $ 909,000.- in five months by predicting other's expectations and speculating "successfully on the basis of this prediction" (Orosel (1985a), p. 15). Similar to the example in 4.1 the reason for this success was the authors possibility to make transactions before (!) other investors follow him.

5. Conclusion

This paper investigated a new view of estimating the present value of shares using the expectations of other investors on future market prices. The following thesis are derivable from the theoretical results.

(T1) Investors should not only spend money for information on the future performance of corporations, but also for information on the subjective expectations and strategies of other investors.

With heterogeneous expectations the fundamental analysis of the earning power of a share is not sufficient to describe its trend of prices. This becomes evident by the ratings of not too big well known firms with growth due to conditions in the particular trade like 'Boss', Stuttgart.

(T2) Portfolio investments must be managed constantly even when all relevant price schemes are free of arbitrage.

This thesis is also a consequence of heterogeneous expectations: Contrary to the portfolio theory it was shown in this model that there are states, in which it is recommendable to buy or to sell shares respectively. A buy-and-hold-strategy cannot be optimal.

(T3) Market prices free of arbitrage can be greater than all investors'present values of expected dividends.

Since it is assumed that trading is possible in all future states, prices do not only reflect expectations on future dividends, but also expectations on other investors'expectations on future market prices.

(T4) Investors should try to investigate, whether signalling their expectations could have positive effects on their returns from investments.

First investors would have to prove whether they are able to send signals which will be recognized from other investors. Contrary to (A10) usually the costs of signals cannot be neglected. Therefore signalling can be recommended only when the costs are less than the expected gains. Second, it won't be easy to get the necessary information to be able to calculate the expected gains. If any, then probably only professional investors will be able to follow this strategy.

(T5) Investors should examine constantly whether their subjective expectations are still sustainable.

Usually inadequate distribution functions on future dividend payments result in losses. To determine the success of a strategy the achieved returns can be compared with the market portfolio. But it is not easy to reach these average returns. Using a formal model Schredelseker (1984a) showed that only outstanding informed investors are able to beat the market. Mühlbradt (1978) supports this thesis using a different approach. His empirical tests on the efficiency of the NYSE deduced, that this stock exchange is efficient in the semi-strong form: Investors are not able to beat the market without any insider information.

(T6) Stock prices depend heavily on the heterogeneity of expectations of investors. The more heterogeneous they are, the higher the stock price can be comparatively to the yield by dividends.

The more heterogeneous expectations of investors are, the less the price of a stock depends on expected dividends, and the more it depends on the expectations, how other investors will value this stock in the future.

The performance of the German stock exchange supports this observation: The stocks of corporations of the electricity industry like 'Veba' or 'RWE' usually are dull, because the expectations of many investors on the future earnings of these corporations of many investors are alike. The yield by dividends is high, because there is no inventiveness in these stocks. The yield by dividends of corporations, which are valued differently by many investors like 'Siemens' or 'Boss', is much lower. Investment decisions to buy these kind of stocks do not depend so strongly on the expected dividends, but on the expectation on further increasing stock prices, i.e. on strenghtened future demand by other investors.

References

Chung, K.L. (1974), A Course in Probability Theory, 2nd ed., New York: Academic Press, 1974.

Fama, E.F. (1970), Efficient Capital Markets: A Review of Theory and Empirical Work, in: Journal of Finance 25, 383-417.

Grossman, S.J. (1981), An Introduction to the Theory of Rational Expectations Under Asymmetric Information, in: Review of Economic Studies 48, 541-559.

Harrison, M.J. and D.M. Kreps (1978), Speculative Investor Behavior in a Stock Market With Heterogeneous Expectations, in: Quarterly Journal of Economics 92, 323-336.

Hirshleifer, J. (1971), The Private and Social Value of Information and the Reward to Inventive Activity, in: American Economic Review 61, 561-574.

Jaffe, J.F. (1974), Special Information and Insider Trading, in: Journal of Business 47, 410-428.

Kahneman, D. and A. Tversky (1979), Prospect Theory: An Analysis of Decision Under Risk, in: Econometrica 47, 263-291.

Keynes, J.M. (1936), The General Theory of Employment, Interest and Money, London: MacMillan, 1936.

Kromschröder, B. (1984), Schlägt Dummheit Mittelmäßigkeit? in: Zeitschrift für betriebswirtschaftliche Forschung 36, 732-747.

Mühlbradt, F.W. (1978), Chancen und Risiken der Aktienanlage: Untersuchungen zur 'efficient market' Theory in Deutschland, Köln: Wison, 1978.

Orosel, G.O. (1985a), Stock Prices and Market Efficiency with Heterogeneous Information, Working Paper No. 8507, University of Vienna.

Orosel, G.O. (1985b), Infinite Horizon Rational Expectations Equilibrium in a Competitive Market for an Exhaustible Resource, in: International Economic Review 26, 701-720.

Samuelson, P.A. (1965), Proof That Properly Anticipated Prices Fluctuate Randomly, in: Industrial Management Review 6, 41-49.

Schmidt, R.H. (1976), Aktienkursprognose, Wiesbaden: Gabler, 1976.

Schredelseker, K. (1984a), Anlagestrategie und Informationsnutzen am Aktienmarkt, in: Zeitschrift für betriebswirtschaftliche Forschung 36, 44-59.

Schredelseker, K. (1984b), Dummheit schlägt Mittelmäßigkeit, in: Zeitschrift für betriebswirtschaftliche Forschung 36, 1074-1079.

Schredelseker, K. (1986), Kapitalmarktgleichgewicht und Information, Working Paper No. 95, University of Wuppertal.

Spence, M. (1974), Market Signaling: Informational Transfers in Hiring and Related Screening Processes, Cambridge, MA: Harvard University Press, 1974.

Swoboda, P. (1982), Heterogene Information und Kapitalstruktur der Unternehmung, in: Zeitschrift für betriebswirtschaftliche Forschung 34, 705-727.

Dividend Policy under Asymmetric Information

Thomas Hartmann-Wendels

Summary: In a perfect capital market the firm's dividend policy is irrelevant to its market value. This theoretical result, however, is in sharp contrast to the observed behavior of firms. It is often argued, that dividend policy is used by the firm's management as an instrument to inform other market participants about the firm's expected future earnings. The potential information content of dividends is investigated in a signalling framework. It is shown that a signalling equilibrium can only exist if dividend policy is not irrelevant even without informational effects. Thus a theory of optimal dividend policy under moral hazard is developed as a basis for the construction of dividend signalling models.

Table of Contents:

1. Introduction

Much of the theoretical work about dividend policy was stimulated by the Miller/Modigliani (MM) irrelevance proposition (1961). While on the one hand the MM irrelevance result was extended to more general market

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

settings[1], it was on the other side investigated, how this result changes, if some critical assumptions of the MM-model were relaxed. Assuming that retained earnings are taxed at a lower rate than dividends, it is optimal to pay dividends only to the extent that funds are not needed for investment. Although such a tax system is valid for the USA and up to 1977 also for Germany[2], it happens on the other side that firms are paying dividends while simultaneously selling new shares. Excluding irrational firm behavior this evidence contradicts the theoretical result. Recent research has investigated the problem of dividend irrelevance under two different aspects. Motivated by the work of Jensen and Meckling (1976) the firm was portrayed as a collection of competing groups whose interests can conflict. So it was argued, that dividend policy is involved in the conflict of interests between stockholders and bondholders, because the former have incentives to transfer wealth from the bondholders by paying investment or debt financed dividends.[3] Why this argument may be able to explain the existence of an optimal dividend policy, no clearcut results have yet been derived.

Under the key-word "signalling-theory" other authors have tried to deal with the thesis that dividend policy conveys information about the management's expectations of future cash flows. While this argument has been known for a long time, it never had been investigated theoretically up to some years ago. Already MM (1961) mentioned, that dividend policy may not be a matter of irrelevance, if a change in the payout ratio is interpreted as a message, that management anticipates permanently higher levels of cash flows from investments. This presumption is based on Lintner's empirical work (1956), which showed that a lot of firms exhibit a behavior of constant dividend payouts. Dividends are only increased, if management is relatively certain to maintain the higher dividend level in the future. Given this type of management behavior, out-

---

1 Franke (1971), Hirshleifer (1974), Stiglitz (1974), Hax (1982), Haley/Schall (1979);

2 In 1977 the so called "Anrechnungsverfahren" was introduced in German tax law: According to this procedure taxes paid on dividends by the firm are credited against the personal income tax of the shareholders. Ignoring flotation costs dividends are preferred to retained earnings by all stockholders, whose personal income tax rate is below the corporate income tax rate. For detailed computations including flotation costs see Hax (1979).

3 See Kalay (1979), (1981), (1982) and Smith/Warner (1979).

side investors will interpret an increase in current dividend payout as a signal that management expects permanently higher levels of cash flows.

A lot of empirical work dealed with the information content hypothesis, but the results are mixed: While some authors had to refute this hypothesis[4], others came to the conclusion that dividends do convey some information about future earnings.[5]

Applying the signalling theory, developed by Spence (1973, 1974), to financial markets, Ross (1977, 1978) was the first, who examined financial decisions as a signalling device. Others followed his approach and explained the suggested information content of dividends within a signalling framework. A characteristic feature of the signalling theory is the condition that information is regarded as authentic, only if the better informed individual has no incentive to signal fraudulently. To provide the appropriate incentives paying dividends must induce costs and/or benefits. Thus this paper concentrates on the necessity to explain the existence of dividend related costs and benefits. It is shown that the existence of an optimal dividend policy is a necessary, but not sufficient precondition for the construction of dividend signalling models. Although it is far beyond the scope of this paper to develop a comprehensive theory of corporate dividend policy, some contributions are made to the relevance of dividend policy in a world of asymmetric information and opportunistic behavior. Subsequently it is investigated how the results of such a theory can be used to explain the information content of dividends within a signalling framework.

This paper is organized in the following manner: In section two the basic concepts of signalling theory are introduced. Section three deals with the existence of an optimal dividend policy. First it is shown that tax arguments do not provide us with the appropriate cost structure signalling theory requires. A review of the existing signalling models yields that the existence of an optimal dividend policy there is based on ad hoc arguments rather than on a well-founded theory. A theory of corporate dividend policy, which is based on conflicts of interests between stockholders and bondholders is developed in the second part of

---

4 Watts (1973), Ang (1975), Gonedes (1978);

5 Pettit (1972), (1976), Laub (1976), Aharony/Swary (1980), Brickley (1983), Patell/Wolfson (1984), and for Germany: Brandi (1977), Sahling (1981);

section three. This is accomplished by extending Kalay's (1979, 1982) approaches concerning dividend constraints. The signalling effects related to the existence of an optimal dividend policy are discussed in section four. Section five provides a summary and some conclusions for further research.

2. The Basic Concepts of Signalling Theory

Signalling theory applies to markets, where asymmetric information prevails about the quality of a product. In the case of dividend policy it is presumed that the firm's management is better informed about the firm's future cash flows than outside capital market investors. Dividend policy then serves as a signal that conveys information about these expectations. Outside investors will regard a signal as trustworthy, if they know that the management has no incentive to signal falsely. Such is the case, if it is too costly for a management to imitate the signalling behavior of a firm with a higher level of expected future cash flows. Therefore it is an essential feature of signalling theory that signalling induces costs and/or benefits[6], which are related to the unknown true quality of the product, i.e. paying dividends must lead to costs and/or benefits, which are related to a firm's expected earnings.

If the dividend policy is regarded as a signalling device, the market price of a firm's shares, reflecting the expectations of the outside investors, depends on the amount of dividend payment (D):

$$P = P(D) \qquad \frac{d\,P(D)}{d\,D} > 0 \tag{1}$$

An increase in market price provides benefits to the stockholders. These benefits may also depend on the "true" level of the unknown product quality, i.e. on the level of future cash flows, as they are expected by the better informed management. The level of expected "true" cash flows is parameterized by θ. Thus we have for the signalling (gross-) benefits:

6 Within this paper we will only deal with dissipative signalling structures.

$$G = G\big(P(D),\theta\big)$$

$$\frac{\partial G\big(P(D),\theta\big)}{\partial P} > 0 \tag{2}$$

In order to establish a signalling equilibrium we have to introduce signalling costs, which must depend on both, dividend payments and θ, the unknown parameter:

$$C = C(D,\theta)$$

$$\frac{\partial C(D,\theta)}{\partial D} > 0 \tag{3}$$

The management now determines the optimal signal, i.e. the dividend payment, that maximizes the difference of signalling benefits and signalling costs. Of course this optimum depends on the equilibrium relationship between P and D. Every firm $j \in J$ is characterized by the level of expected true cash flow θ_j, and $\{D_{jk} | k \in K\}$ is the set of all possible signals for firm type j. Then we have the following first part of a definition of a signalling equilibrium:

Every manager of a firm $j \in J$ selects the optimal signal (i.e. the optimal dividend payment) D_j^*:

$$G\big(P(D_j^*),\theta_j\big) - C(D_j^*,\theta_j) > G\big(P(D_{jk}),\theta_j\big) - C(D_{jk},\theta_j) \tag{4}$$

for all $k \in K$, except where $D_{jk} = D_j^*$ for all $j \in J$.

The second part of the equilibrium definition requires, that the expectations of the capital market investors are fulfilled, i.e. that the market price of the firm's shares P(D) equals its (fictive) "true" market value V(θ,D). This "true" market value is the market price, which would prevail if all investors had the same information the management has. As this true market value may also depend on the level of the signal, we have

$$\frac{\partial V(\theta,D)}{\partial \theta} > 0 \quad \text{and} \quad \frac{\partial V(\theta,D)}{\partial D} \gtreqless 0 \tag{5}$$

The second part of the equilibrium definition then is:

$$P(D_j^*) = V(\theta_j, D_j^*) \qquad \text{for all} \quad j \in J \tag{6}$$

As can be seen from (4) the manager's optimization problem includes P(D), whose value is determined by the expectations of the outside investors. In a signalling equilibrium the expectations of the outside investors are such, that they induce a signalling device of the better informed, that in turn fulfills these original expectations. Thus the expectations have the feature of a self-fulfilling prophecy, a feature the signalling models share with some rational expectations models. And indeed, the signalling equilibrium can be interpreted as a specific type of rational expectations equilibrium.[7] As it is known from rational expectations theory, there exists an infinite number of solutions for the equilibrium relationship $D = f(\theta)$, which satisfy (4) and (6). In order to yield a unique solution, it is necessary to introduce an additional boundary condition. So it is often requested in signalling models, that the one with the lowest level of θ does not signal, because he cannot reap any benefits from signalling.

There are two necessary and sufficient conditions for a signalling equilibrium to exist.[8] The first one requires the existence of an optimal D, even without signalling effects, within an interval $(0,\overline{D})$:

For every θ_j, $j \in J$, there exists a $\overline{D}_j \geqq 0$, such that for all $D_j > \overline{D}_j$

$$\frac{\partial C(D,\theta)}{\partial D} - \frac{\partial G(P(D),\theta)}{\partial P}\,\frac{\partial V(\theta,D)}{\partial D} > 0 \tag{7}$$

where $\overline{D}_j$ is defined by the following conditions:

if $\frac{\partial V(\theta,D)}{\partial D} \leqq 0$: $\overline{D}_j = 0$

if $\frac{\partial V(\theta,D)}{\partial D} > 0$: $\overline{D}_j$ is that value of the signal, that fulfills (7) as an equality.

7 This point is worked out more detailed in Hartmann-Wendels (1986).

8 See Riley (1979) for a detailed proof.

$\overline{D}_j$ is therefore by definition the level of the signal, that would be optimal for firm type θ_j, if D did not convey any information. Suppose (7) is not fulfilled, then everyone will invest in D up to the highest possible value, such that D cannot convey any information.

The second part of the equilibrium condition is more restrictive and refers to the relationship between signalling costs and signalling benefits:

$$\frac{\partial}{\partial \theta}\left(\frac{\dfrac{\partial C(D,\theta)}{\partial D}}{\dfrac{\partial G\left(P(D),\theta\right)}{\partial P}}\right) < 0 \qquad \text{for all } \theta_j,\ j \in J \tag{8}$$

The numerator can be interpreted as the marginal signalling costs, and the denominator is the marginal benefit from an increase in market price. Thus (8) requires in essence, that the relationship between marginal signalling costs and marginal signalling benefits is negatively related to the value of θ, i.e. to the level of expected cash flows.

In the case of signalling benefits independent of θ, (8) reduces to:

$$\frac{\partial^2 C(D,\theta)}{\partial D\ \partial \theta} < 0 \tag{9}$$

Condition (9) is the known Spence-criterion, that marginal signalling costs must be negatively related to θ.[9] The conditions (8) or (9) respectively require, that it must be unprofitable for a low θ-type to imitate the signalling decisions of a firm with a higher θ.

In order to apply signalling theory to the thesis of the information content of dividend policy, it has to be explained, why costs and benefits may be connected with dividend payouts. If such costs and benefits do exist, then there must also exist an optimal dividend policy even in the absence of signalling effects. Dividend policy, therefore, can only be regarded as a signalling device, if dividend policy is not irrelevant for the firm's market value, even without signalling. This is the reason, why a dividend signalling model cannot be a simple extension

---

9 Spence (1973), (1974).

of the MM irrelevance proposition, because a precondition for a signalling equilibrium is, that the irrelevance proposition is not valid.

Whereas signalling theory cannot explain the existence of an optimal dividend policy, signalling effects have an impact on the optimal dividend payout, provided an optimal dividend policy exists without signalling. In the absence of signalling-effects an optimal dividend policy in general can be determined by the first order condition:

$$\frac{\partial C(D,\theta)}{\partial D} = \frac{\partial G\left[P(D),\theta\right]}{\partial P} \frac{\partial V(\theta,D)}{\partial D} \tag{10}$$

In the optimum, the marginal costs of dividend payments equal the marginal benefits of a rise in firm value. The nature of the costs and/or benefits, associated with dividend payments, remain to be specified of course.

In the signalling case it must be borne in mind, that a rise in market value may not only stem from a dividend payment directly, but also indirectly via the impact of a dividend payment on the market assessed value of θ. Therefore the first order condition for the optimal dividend payout in the signalling case is:

$$\frac{\partial G\left[P(D),\theta\right]}{\partial P} \frac{\partial V(\theta,D)}{\partial \theta} \frac{d\theta}{dD} + \frac{\partial G\left[P(D),\theta\right]}{\partial P} \frac{\partial V(\theta,D)}{\partial D} = \frac{\partial C(D,\theta)}{\partial D} \tag{11}$$

From (11) it is obvious that the optimal dividend payment in the signalling case is higher than in the no-signalling case. In a welfare economic context this result was denoted as the overproduction feature of signalling equilibria.[10]

10 Spence (1974). The overproduction feature of signalling equilibria gives rise to instability problems. Riley (1979).

3. Towards a Theory of Corporate Dividend Policy

3.1. Survey of Existing Dividend Signalling Models

Applying signalling theory to dividend policy we are now left with the result, that we first have to explain why dividend policy is not irrelevant in a world without signalling effects.

A first reason, why dividend policy may not be irrelevant for the market value of a firm is the existence of taxes. If dividends are taxed at a higher rate than retained earnings, it is optimal to pay no dividends. Thus using dividends as a signalling device imposes costs, but unfortunately these costs do not fulfill condition (8), i.e. they are not negatively related to the value of θ, as long as the tax rate rate is independent of the level of earnings. But as Bhattacharya (1979) has shown, the existence of taxes has an impact on the equilibrium relationship between D and θ. Given some other signalling costs with the appropriate cost structure, the optimal level of D in an equilibrium will be lower for every θ, if there exists a tax regime which discriminates dividend payments.

In order to explain the existence of an optimal dividend policy, we may look, how the existing dividend signalling models resolve this problem. Bhattacharya (1979) assumes, that there is no frictionless access to the capital market, if a firm needs outside capital, because the cash flow does not suffice to pay the previously announced dividend. To make up the shortage, the firm has to accept a higher interest rate or to postpone or even cancel planned investments. But there is no clear justification for this penal interest rate, and it is doubtful, whether it can survive in a competitive credit market.

Applying the incentive-signalling approach of Ross (1977, 1978) to dividend policy Kalay (1979, 1980) assumes, that the manager's compensation is tied to the market price of the firm's shares, which in turn is an increasing function of the dividend payout. Furthermore the manager incurs a penalty if he cuts the dividend payment. Therefore the manager would never shorten the dividend payment unless he is forced to do so. In order to provide the manager with the proper incentives to signal the "true" firm value, we have to put some restrictions on dividend payments subject to the firm's cash flow. This rises the question, how we can justify such a dividend restriction. Why should it not be possible to raise outside capital in order to make up a financial distress caused by a dividend payment higher than cash flows?

The same problem arises in Heinkel's (1978) dividend signalling model. There it is assumed, that only shares can be issued and that the amount of new equity is restricted to the difference between present investment and retained earnings of the preceding period. Thus if the firm wants to pay dividends, it has to cut its investment program to the same extent. The Heinkel-model has the remarkable feature, that signalling costs are positively related to the firm's cash flow: a reduction of the investment program is more costly the higher the profitability of the cancelled investment projects is. Nevertheless the equilibrium condition (8) is satisfied, because the signalling benefits are also positively related with the future cash flow: The higher the future cash flows are assessed by the market investors, the higher is the price, at which the new shares can be issued, and the higher is the fraction that remains for the old shareholders. The value of this remaining fraction, in turn, is the higher, the higher are the firm's "true" expected cash flows. This is the reason, why the Heinkel-model cannot be extended to the case of debt financing. If the outstanding bonds are risky, the required rate of return depends only on the assessed future cash flows but not on the "true" expected future cash flows. Thus there is no possibility for firms with higher expected cash flows to keep firms with lower expected cash flows from imitating their signalling decisions.

Miller and Rock (1985) avoid the problem of justifying a dividend restriction, in that they consider the dividend payment net off any new equity or debt financing. Because the capital market investors are able to calculate the optimal level of new investment, the net-dividend payment provides them with the "missing piece"[11], in order to determine the level of present cash flow. As present and future earnings are assumed to be positively correlated, the present dividend policy also informs about the future earnings of a firm. Although the Miller/Rock-model is not a proper signalling model but a rational expectations model, it can be interpreted as a signalling model. While Miller and Rock analyse the information content of dividend policy under the assumption that a rational expectations equilibrium does exist, we have to ask in a signalling framework, whether such an equilibrium can survive in a world, where managers set false signals if they have incentives to do

11 Miller/Rock (1985), p. 11.

so. Only if we can show that such adverse incentives do not exist, i.e. if condition (8) holds, then the rational expectations equilibrium is also a signalling equilibrium. The costs of a dividend payout in the Miller/Rock-model stem from a reduction of investments below the value maximizing level. In order to pay a certain amount of dividends, this reduction must be higher for firms with low present cash flows, so that signalling costs are inversely related to present earnings, which in turn are an indicator for future earnings. Therefore the Miller/Rock-model can also be regarded as a signalling model.

The preceding analysis shows, that we have to explain, why dividend payments are bounded by the firm's earnings. Otherwise it would always be possible to pay a large amount of dividends by raising new capital through the issuance of new shares or bonds at the same time, without incurring any loss in market value. In such a world the dividend policy could not be used as an informational instrument with respect to the future prospects of a firm.

It is well known, that financing decisions may not be a matter of irrelevance, if we do not assume firm value maximizing behavior, but allow agents to pursue their own interests at the expense of others. Especially dividend policy can be used to shift wealth from the bondholders to the stockholders, if the outstanding bonds are risky, and the dividend payment is financed by a new debt issue or reduced investment. In order to avoid such a wealth transfer, stockholders and bondholders have incentives to impose constraints on the firm's ability to pay debt or investment financed dividends.

The nature of the typical dividend constraint is examined in Kalay (1979, 1982). There it is stated, that the dividend constraint is in essence a minimum investment restriction and in John and Kalay (1982) the impact of such a restriction on the investment behavior of a levered firm is examined for a given amount of outstanding debt. But it is evident that the debt induced dividend constraint will influence the firm's initial decision to issue bonds. As a consequence investment and financing decisions are no longer independent but have to be determined simultaneously.

As we have to explain the existence of an optimal dividend policy in order to construct a dividend signalling model, it is not sufficient to prove the mere existence of such an optimum, but, in addition, we must investigate, how this optimum can be determined and how it depends on the unknown parameter, i. e. the expected future earnings. Only if it is possible to specify this relationship, we can state, whether a deviation from this optimum for signalling purposes will induce costs with the proper cost structure, i.e. signalling costs, which fulfill condition (8) or (9) respectively.

In the following section we will determine an optimal dividend policy in a world of moral hazard. As we are only concerned with debt and investment financed dividends, the optimal dividend policy is specified only net off the proceeds of a new equity issue. Otherwise we, in addition, had to examine the firm's ability to pay equity financed dividends, a task, we will not pursue within this paper.

3.2. Corporate Dividend Policy in a Moral Hazard Scenario

3.2.1. Investment Decisions and Debt Financing

In this section a one period model of a levered firm's investment behavior is developed. At time 0 the firm invests an arbitrarily divisible amount ,I, in an investment opportunity yielding revenues at the end of the first period according to the stochastic production function $\theta\sqrt{I}(1+\tilde{\varepsilon})$. The parameter θ measures the firm's profitability and $\tilde{\varepsilon}$ is a random variable, which is equally distributed over the interval (-a,a). In order to restrict the conflict of interests to stock- and bondholders, we assume, $a \leqq 1$. Thus the investment cannot result in a loss that surmounts the initial investment.

The strong concavity of the production function implies that the firm cannot invest arbitrary amounts at the market rate of interests. This opportunity is not feasible for the firm, if the revenues of an investment are taxed at a higher rate for firms than for private investors.

At t = 0 the firm has raised an amount of equity S and of debt B. At the end of the first period the firm has to pay back principal and interests, $B(1 + r^b)$. The interest rate of debt is r^b and for equity r^e, which are assumed to be exogenously given for the moment. For the ease of analytical tractability we furthermore assume risk neutral stockholders and bondholders. At t = 0 the management, which acts by assumption in the stockholders' interest decides how to devide the available amount of funds upon investment or consumption, i.e. dividends. The investment program, which maximizes the stockholders' wealth is determined by

$$\max_{I} \quad Z = -I + B + S + \frac{1}{1+r^e} E\{\theta\sqrt{I}(1+\tilde{\varepsilon}) - \operatorname{Min}\{\theta\sqrt{I}(1+\tilde{\varepsilon});\ B(1+r^b)\}\} \tag{12}$$

where E{•} is the stockholders' expected liquidation revenue at t = 1.

As the production function is distributed in a finite interval, we have to distinguish three different cases:

$$\max_{I} \quad Z = \begin{cases} (1) \quad -I + B + S + \dfrac{1}{1+r^e}\left(\theta\sqrt{I} - B(1+r^b)\right) \\ \qquad \text{if } \theta\sqrt{I}(1-a) \geqq B(1+r^b) \\ (2) \quad -I + B + S + \dfrac{1}{1+r^e}\,\dfrac{\theta\sqrt{I}}{4a}\left(1 + a - \dfrac{B(1+r^b)}{\theta\sqrt{I}}\right)^2 \\ \qquad \text{if } \theta\sqrt{I}(1-a) < B(1+r^b) \leqq \theta\sqrt{I}(1+a) \\ (3) \quad -I + B + S \\ \qquad \text{if } \theta\sqrt{I}(1+a) < B(1+r^b) \end{cases} \tag{13}$$

The term (1) is relevant for such values of I and $B(1+r^b)$, that the credit is riskless. The stockholders' wealth maximizing investment program (I^*_S) then is:

$$I^*_{S(1)} = \frac{\theta^2}{4(1+r^e)^2} = I^* \tag{14}$$

It is well known that there exists no conflict of interests between stockholders and bondholders if the outstanding bonds are riskless, because the stockholders then have to bear the whole risk of the investment alone. Therefore $I^*_{S(1)}$ equals the firm value maximizing investment program I*.

The expression (13 - 2) applies to cases, where the outstanding bonds are risky, but there is still a positive probability that the whole debt may be paid off. The first order condition for the stockholders' wealth maximizing investment program $I^*_{S(2)}$ is

$$1 = \frac{1}{1+r^e}\,\frac{1}{8a\sqrt{I}}\left(\theta(1+a)^2 - \frac{\left(B(1+r^b)\right)^2}{\theta I}\right) \tag{15}$$

and the second order condition

$$\frac{\partial^2 Z}{\partial I^2} = \frac{1}{2\theta I^2}\,\frac{1}{1+r^e}\,\frac{1}{8a\sqrt{I}}\left(I\left(B(1+r^b)\right)^2 - \theta^2 I(1+a)^2\right) < 0 \tag{16}$$

is fulfilled for $\theta^2(1+a)^2 > \left(B(1+r^b)\right)^2$.

Furthermore the relationship $\theta^*_{S(2)} < \theta^*$ is valid. The investment volume that maximizes shareholders' wealth is smaller than the firm value maximizing investment volume, if the firm has risky bonds outstanding.[12]

From (15) it is evident that $I^*_{S(2)}$ now depends on $B(1+r^b)$. Differentiating with respect to $B(1+r^b)$ we get

$$\frac{dI}{d\left(B(1+r^b)\right)} = \frac{4\,B\,(1+r^b) - I}{I\left(B(1+r^b)\right)^2 - \theta^2 I(1+a)^2} < 0 \tag{17}$$

for $4B(1+r^b) > I$, and the denominator is by assumption negative.

From (17) we see, that $I^*_{S(2)}$ is the smaller, the higher the principal and interest payment is.

Differentiating with respect to θ

$$\frac{dI}{d\theta} = -\,\frac{\theta^2 I(1+a)^2 + \left(B(1+r^b)\right)^2}{\theta\left(\left(B(1+r^b)\right)^2 - \theta^2(1+a)^2\right)} > 0 \tag{18}$$

shows us that $I^*_{S(2)}$ is increasing in the profitability parameter .

If it is impossible to pay back interest and principal completely,

---

12 For a detailed investigation and formal analysis of this issue see Hartmann-Wendels (1986). The underinvestment problem is also investigated by Myers (1977).

(13-3) is the relevant part of the stockholders' objective function. Because

$$\frac{\partial Z}{\partial I} = -1 \tag{19}$$

it is evident that nothing will be invested in this case, because it would only reduce the dividend payout at t=0 without increasing the liquidation value at t=1.

3.2.2. Dividend Constraints and Stockholder - Bondholder Conflict

Up to this point we have investigated the investment behavior of a levered firm for a given amount of outstanding debt and a given nominal rate of interest. Rational bondholders will attribute to the management those actions which are in the stockholders' interest. Bondholders, therefore, will pay into the firm in accord with these forecasts. Hence, as it is shown in Jensen/Meckling (1976), it is the shareholders, who have to bear the consequences of any expected wealth transfer, and it is in their interest to constrain their ability to pay investment or debt financed dividends. So we will now investigate the impact of a dividend constraint on the investment behavior of a levered firm.

Kalay (1979, 1982) examines bond indentures and concludes that the typical dividend restriction has the form:

$$D_t \leqq \text{Max } \{0, \sum_{\tau=0}^{t} S_\tau + \gamma \sum_{\tau=0}^{t} G_\tau + F - \sum_{\tau=0}^{t-1} D_\tau\} \tag{20}$$

G_τ : earnings of period τ
γ, F: constants

The most important features of this dividend restriction are:

- The firm is not forced to pay a "negative dividend", i.e. to sell new shares.
- The dividend constraint is cumulative, so that the firm's ability to pay dividends depends on the dividends paid in preceding periods.
- Only investment and debt financed dividends are constrained, whereas proceeds from the sale of new shares and to a certain degree retained

earnings are allowed to be paid out.

- The dividend restriction is in essence a minimum investment commitment. In order to see this we apply the cash flow identity:

$$G_\tau + Dep_\tau + L_\tau + S_\tau + B_\tau = D_\tau + I_\tau + P_\tau \quad (21)$$

The left hand side represents the inflows, consisting of earnings (G_τ), depreciation (Dep_τ), proceeds from the sale of assets (L_τ), shares (S_τ), and bonds (B_τ). The right hand side represents the outflows of a period as the sum of (gross) dividends (D_τ), investment (I_τ), and repayment of debt (P_τ). Cumulating (21) over t periods and substituting in (20) for D_τ, we get the minimum investment commitment for the case that the firm is able to pay a positive dividend:

$$\sum_{\tau=0}^{t} (I_\tau - Dep_\tau - L_\tau) \geqq \sum_{\tau=0}^{t} ((1-\gamma) G_\tau + B_\tau - P_\tau) - F \quad (22)$$

The left hand side of (22) is the cumulative net investment, which must surmount the cumulative sum consisting of that portion of the earnings that has to be retained and of the net proceeds of new bonds less the initial fund F.

In the one period case under the assumption, $\gamma=1$, (22) reduces to:

$$I_t \geqq B_t - F \quad (23)$$

For the ease of tractability we will use (23) in the form [13]:

$$B_t \leqq \alpha I_t \quad (24)$$

3.2.3. Optimal Investment Policy under a Dividend Constraint

For a given amount of debt, John and Kalay (1982) examine the impact of a dividend or minimum investment restriction, respectively, on the corporate investment behavior. But if the issuance of new debt imposes a

13 In Kalay's sample γ has the value 1 for 106 out of 128 bond indentures, and the mean was 0.95. On average 68.2% (=1.47) of the amount of debt financing had to be invested. Kalay (1982), p. 217-218

minimum investment restriction on the firm, the stockholders will take this into consideration when they decide upon debt financing. Thus the optimal investment and debt financing level has to be determined simultaneously.

At this point it is necessary to explain, why a firm issues debt although this induces adverse incentives which impose agency costs on the stockholders. In order to rationalize debt financing it will be assumed that the costs of debt financing are lower than of equity financing, probably because interests are tax deductable. The bondholders, knowing this, expect the firm to raise as much debt as is possible for a given amount of planed investment so that the firm's feasibility to issue bonds will be completely exhausted. Furthermore we will assume that debt financing is restricted such that the complete repayment of principal and interests never becomes impossible.

Given risk neutral behavior and assuming a competitive bond market the bondholders will require a nominal interest rate $\left(1+r^{b}(B)\right)$, such that the firm's expected payment of principal and interests yields a return at the risk free rate (r^{bs}):

$$B(1 + r^{bs}) = \begin{cases} B(1 + r^{bs}) \\ \text{if } \theta\sqrt{T}(1-a) > B\left(1 + r^{b}(B)\right) \\ \\ B\left(1 + r^{b}(B)\right) \int_{\beta}^{a} \frac{1}{2a} \, d\varepsilon + \frac{1}{2a} \int_{-a}^{\beta} \theta\sqrt{T}(1 + \tilde{\varepsilon}) \, d\varepsilon \\ \\ \text{if } \theta\sqrt{T}(1-a) \leqq B\left(1 + r^{b}(B)\right) \\ \text{and } B(1 + r^{bs}) \leqq \theta\sqrt{T} \end{cases} \qquad (25)$$

$$\beta = \frac{B\left(1 + r^{b}(B)\right)}{\theta\sqrt{T}} - 1$$

Solving (25) for $\left(1 + r^{b}(B)\right)$ shows that the required nominal rate of interest depends on the firm's expected investment behavior:

$$1 + r^b(B) = \begin{cases} 1 + r^{bs} \\ \text{if } \theta\sqrt{T}(1-a) > B\left(1 + r^b(B)\right) \\ \\ \dfrac{\theta\sqrt{T}}{B}(1+a) - 2\sqrt{\dfrac{a\theta\sqrt{T}}{B}\left(\dfrac{\theta\sqrt{T}}{B} - (1 + r^{bs})\right)} \\ \text{if } \theta\sqrt{T}(1-a) \leqq B\left(1 + r^b(B)\right) \\ \text{and } B(1 + r^{bs}) \leqq \theta\sqrt{T} \end{cases} \tag{26}$$

Assuming that it is optimal for the stockholders to exhaust the minimum investment restriction completely, (24) will be fulfilled as an equality and we have:

$$I = \frac{B}{\alpha} \tag{27}$$

Substituting (27) in (26) we can express $\left(1 + r^b(B)\right)$ as a function of the expected investment program:

$$1 + r^b(I) = \begin{cases} 1 + r^{bs} \\ \text{if } \theta\sqrt{T}(1-a) > B\left(1 + r^b(B)\right) \\ \\ \dfrac{\theta}{\alpha\sqrt{T}}(1 + a) - 2\sqrt{\dfrac{a\theta^2}{\alpha^2 I} - \dfrac{a\theta(1 + r^{bs})}{\alpha\sqrt{T}}} \\ \text{if } \theta\sqrt{T}(1-a) \leqq B\left(1 + r^b(B)\right) \\ \text{and } B(1 + r^{bs}) \leqq \theta\sqrt{T} \end{cases} \tag{28}$$

Substituting (27) and (28) in the objective function (12) yields:

$$\max_I \{Z = I(\alpha - 1) + S + \frac{1}{1+r^e}\left(\theta\sqrt{T} - \alpha I(1 + r^{bs})\right)\} \tag{29}$$

$$\text{if } B(1 + r^{bs}) \leqq \theta\sqrt{T}$$

The stockholders' wealth maximizing investment program is

$$I_S^* = \frac{\theta^2}{4\left(\alpha(r^{bs} - r^e) + (1 + r^e)\right)^2} \tag{30}$$

and is independent, whether the outstanding bonds are risky or not. Whether this is the case depends only on the parameter values of a, α, r^e, and r^{bs}, given the investment program I^*_S. The optimal investment program, determined by (30), is not only the stockholders' wealth maximizing investment program, but it is also the firm value maximizing investment volume subject to a dividend constraint. This result expresses the well known fact that the first best solution can be achieved if the conflict of interests can be settled without incurring any transaction costs.[14]

The optimal amount of debt financing then is:

$$B^* = \alpha I^*_S \tag{31}$$

Assuming the outstanding bonds to be risky an amount of debt, $B < \alpha I^*_S$, would transfer wealth from the stockholders to the bondholders, because the riskiness of the bonds is lower than has been assessed by the bondholders at the time of contracting. Furthermore the stockholders would forego some of the tax advantages of debt financing.

The expected dividend of period t, immediately before the realization of $\tilde{\varepsilon}$ is known but after the new investment program has been determined is given by:

$$D^N_t = D_t - S_t = B_t - I_t + E\{\text{Max } 0,\ \theta_{t-1}\sqrt{I}_{t-1}(1+\tilde{\varepsilon}) - B_{t-1}(1+r^b)\} \tag{32}$$

In (32) D^N_t is a dividend net off new equity. Assuming $\theta_{t-1} = \theta_t = \theta$, and inserting (28), (30), and (31) in (32) yields:

$$D^N_t = I^*\left(2\left(\alpha(r^{bs} - r^e) + (1 + r^e)\right) - (1 + r^{bs})\right) \tag{33}$$

As I^* is an increasing function of θ and as the term in brackets is positive, D^N_t is also increasing in the profitability parameter θ.

4. Signalling under a Theory of Optimal Dividend Policy

Using the results of the preceding analysis for the construction of a

---

14 This is the essential content of the Coase-theorem. Coase (1960).

dividend signalling model, we assume that all firms face the same type of production function and that the parameter values of a, α, r^e, and r^{bs} are common knowledge. The firms differ only in the value of θ, which is known only to insiders, i.e. to the management of a firm and for the moment also to the bondholders. A signalling equilibrium is only viable if it is too costly for a firm with a low θ to mimic the dividend payment of a firm with a higher θ. Assuming $\alpha > 1$, the firm has to extend the investment level beyond I* in order to pay a higher dividend.[15]

Substituting $D(1/\alpha - 1)$ for I in (29) and differentiating with respect to D yields:

$$- \frac{1}{1+r^e} \frac{\alpha}{\alpha - 1} (1 + r^{bs}) + 1 + \frac{\theta}{2(1 + r^e)} \sqrt{\frac{1}{(\alpha - 1)D}} = - \frac{\partial C(D,\theta)}{\partial D} \qquad (34)$$

Changing sign, (34) can be interpreted as signalling costs. Differentiating (34) with respect to θ, shows that the cost structure fulfills the equilibrium condition (9):

$$\frac{\partial^2 C(D,\theta)}{\partial D \; \partial \theta} = - \frac{1}{2(1+r^e)} \sqrt{\frac{1}{(\alpha - 1)D}} < 0 \qquad (35)$$

While in the no-signalling case the optimal dividend policy is determined by the condition that (34) - i.e. marginal signalling costs - equals zero, this condition is replaced in the signalling case by the requirement that marginal signalling costs equal marginal signalling benefits. Without detailed investigation of these benefits it is obvious that the optimal dividend, net off new equity, in the signalling case is higher than without signalling.

In order to determine the exact value of the optimal dividend in the signalling case we have to specify the signalling benefits. These may result from several sources:

- First we may consider the case that a fraction of the firm's stockholders wants to sell their shares. Assuming for simplicity that the stockholders can be divided into one group that wants to hold all their shares and into another group that wants to sell all their shares, the firm's management is confronted with a conflict of interests among their

---

15 In the case $\alpha < 1$, a similar argument as with respect to the Miller/ Rock-model does apply.

stockholders:[16] The selling group prefers the dividend policy that signals the highest possible market value, irrespectively of any signalling costs, because these stockholders are only interested in the current market price of their shares. The holding group however has to bear the costs of any deviation from the optimum according to (34) without taking any benefit from an increase in current market price. A compromise could take the form that both groups are taken into account according to their relative weights. But then the optimal dividend policy in the signalling case depends upon the numerical values of this weights, so that there is generally no one-to-one correspondence between D and θ.

- The market price of shares is also relevant for stockholders' wealth if the firm is going to issue new shares. Given the amount of new equity the portion that remains to the old stockholders is the larger, the higher the market price of the new shares is. As Heinkel (1978) has shown, the benefits of this remaining portion are increasing in the "true" market value of the firm, so that the signalling benefits are positively correlated to θ.

- Signalling effects are also relevant in the case of debt financing. Assuming the outstanding bonds to be risky, signalling higher returns reduces the required (nominal) interest rate and makes it profitable to invest more. While the benefits of a reduced interest rate are independent of the "true" value of θ, the returns of a larger amount of investment in the signalling case are increasing in θ. Thus we have in this scenario, in contrast to Heinkel's model, also signalling benefits, that are positively related to θ in the case of debt financing.

Furthermore we can apply our model of optimal dividend policy to the incentive signalling approach. In order to accomplish this we first have to invoke the "reluctance-to-cut-dividends" hypothesis. Therefore we assume, that the manager incurs a penalty in the event of a dividend reduction. On the other side his compensation is (at least partially) increasing in the dividend payout, so that the management has incentives to pay high dividends.

Second we must explain why the manager is forced to cut dividends if he signals fraudulently. The highest feasible amount of dividend payout at the beginning of period t depends on the investment and financing

---

16 This kind of conflicting interests are investigated by Miller/Rock.

projects undertaken in t-1 as well as on the event, which prevails at t. On the other side the new investment and financing program starting at t also determines the amount that can be paid out as dividends at the beginning of period t. We assume that $\tilde{\varepsilon}(s_t)$ is an intertemporally independently and identically distributed random variable. Furthermore $\tilde{\theta}_t$ is a random variable depending on the entrance of event s_t at the beginning of period t. There exist stochastic dependencies between the distributions of $\tilde{\theta}(s_t)$ at different periods and $E_t(\tilde{\theta}_{t+h}/s_t)$ denotes the expectation of θ at period t+h under the condition that s_t has occurred.

In this scenario it can be optimal for a manager to pay a lower dividend than the highest possible one. This is the case, if an event s_t occurs such, that $\varepsilon(s_t)$ has a high value, whereas $\theta(s_t)$ indicates only poor investment opportunities so that a current high dividend payment cannot be sustained in future periods. As the management's compensation depends on the dividend level as well as on a reduction of dividend payments, the manager takes into consideration not only the current opportunities to pay dividends but also the future expected amount of payable funds.

Within this model there is also an explanation for the thesis that dividend announcements signal the long run expectations of a firm's management. To show this, we assume, that the occurrence of event s_t leads to a change of all $E_t(\tilde{\theta}_{t+h}/s_t)$, $h > 0$, into the same direction. Then the intertemporally independently and identically distributed $\tilde{\varepsilon}(s_t)$ represent the short run oscillations of earnings, whereas $E_t(\tilde{\theta}_{t+h}/s_t)$ indicates the long run expectations. If we furthermore assume that the manager's compensation is such that it depends more on the dividend reduction component than on the dividend level component, dividend decisions are predominantly influenced by long run expectations about future earnings.

5. Summary and Conclusions

Applying signalling theory to dividend policy makes it necessary to explain, why dividend policy is not a matter of irrelevance even in the absence of signalling effects. A survey of the existing dividend signalling models yields that there the existence of an optimal dividend policy is the result of some ad hoc assumptions, which are difficult to rationalize.[17] Therefore the purpose of this paper was to contribute to

---

17 This criticism extends to other financial signalling models, which

the development of a theory of optimal dividend policy, which rests upon moral hazard arguments. In order to apply the results of this analysis to the construction of dividend signalling models we needed a detailed specification of the optimal dividend policy and its influencing factors. For this reason we choose a one-period model, a specific production function, as well as a specific form of dividend constraint. A richer theory of optimal dividend policy, of course, must include multi-period problems as well as several bond covenants. Furthermore such a theory must investigate the alternatives of internal versus external equity financing. As this problem was ignored within this paper our results apply only to dividends net off new equity.

Finally it was investigated how the theory of optimal dividend policy, developed in section 3, can be used in a signalling framework. While this investigation provides only little formal analysis, it should be evident that the explanation of the information content of dividends in a signalling framework may be feasible if we have a well developed theory of corporate dividend policy. Thus we can conclude that the main problem in constructing a dividend signalling model is the development of a theory of corporate dividend policy. But much remains to be done in this area.

References

Aharony, J./Swary, I. (1980): Quarterly Dividend and Earnings Announcements and Stockholders' Returns: An Empirical Analysis, Journal of Finance, Vol. 35, pp. 1-12.

Ang, J.S. (1975): Dividend Policy: Informational Content or Partial Adjustments?, Review of Economics and Statistics, Vol. 57, pp. 65-70

Bhattacharya, S. (1979): Imperfect Information, Dividend Policy, and the Bird in the Hand Fallacy, Bell Journal of Economics, Vol. 10, pp. 259 -270.

Brandi, E.H. (1977): Informationswirkungen der Jahresabschlußveröffentlichung auf Entscheidungen am Aktienmarkt, Doctoral Dissertation, University of Augsburg.

Brickley, J.A. (1983): Shareholder Wealth, Information Signaling and the Specially Designated Dividend. An Empirical Study, Journal of Financial Economics, Vol. 12, pp. 187-209.

---

are not dividend related. Exceptions are the recent articles of Darrough/Stoughton (1986) and Trueman (1986), who employ the debt/equity ratio and the investment volume, respectively as a signal.

Coase, R.H. (1960): The Problem of Social Cost, Journal of Law and Economics, Vol. 3, pp. 1-44.

Darrough, M.N./Stoughton, N.M. (1986): Moral Hazard and Adverse Selection: The Question of Financial Structure, Journal of Finance, Vol. 41, pp. 501-513.

Franke, G. (1971): Verschuldungs- und Ausschüttungspolitik im Licht der Portefeuille-Theorie, Köln et al.

Gonedes, N.J. (1978): Corporate Signalling, External Accounting, and Capital Market Equilibrium: Evidence on Dividends, Income, and Extraordinary Items, Journal of Accounting Research, Vol. 16, pp. 26-29

Haley, C.W./Schall, L.D. (1979): The Theory of Financial Decisions, 2. ed., New York.

Hartmann-Wendels, T. (1986): Dividendenpolitik bei asymmetrischer Informationsverteilung, Wiesbaden.

Hax, H. (1979): Zur Bedeutung der Körperschaftsteuerreform von 1977 für die Dividendenpolitik von Aktiengesellschaften, Zeitschrift für betriebswirtschaftliche Forschung, Vol. 31, pp. 322-334.

Hax, H. (1982): Finanzierungs- und Investitionstheorie, in: H. Koch et al. (Eds.): Neuere Entwicklungen in der Unternehmenstheorie, Wiesbaden, pp. 49-68.

Heinkel, R. (1978): Dividend Policy as a Signal of Firm Value, from: Essays on Financial Markets with Asymmetric Information, unpubl. diss., University of California, Berkeley.

Hirshleifer, J. (1974): Kapitaltheorie, Köln.

Jensen, M./Meckling, W.H. (1976): Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure, Journal of Financial Economics, Vol. 3, pp. 305-360.

John, K./Kalay, A. (1982): Costly Contracting and Optimal Payout Constraints, Journal of Finance, Vol. 37, pp. 457-470.

Kalay, A. (1979): Corporate Dividend Policy: A Collection of Related Essays, unpubl. diss., University of Rochester.

Kalay, A. (1980): Signalling, Information Content, and the Reluctance to Cut Dividends, Journal of Financial and Quantitative Analysis, Vol. 15, pp. 855-865.

Kalay, A. (1981): Towards a Theory of Corporate Dividend Policy, unpubl. manuscript, University of New York.

Kalay, A. (1982): Stockholder - Bondholder Conflict and Dividend Constraints, Journal of Financial Economics, Vol. 10, pp. 211-233.

Laub, M. (1976): On the Informational Contents of Dividends, Journal of Business, Vol. 49, pp. 73-80.

Lintner, J. (1956): Distribution of Incomes of Corporations among Dividends, Retained Earnings, and Taxes, American Economic Review, Vol. 32, pp. 97-113.

Miller, M.H./Modigliani, F. (1961): Dividend Policy, Growth and the Valuation of Shares, Journal of Business, Vol. 34, pp. 411-433.

Miller, M.H./Rock, K. (1985): Dividend Policy and Asymmetric Information, Journal of Finance, Vol. 40, pp. 1031-1052.

Myers, S.C. (1977): Determinants of Corporate Borrowing, Journal of Financial Economics, Vol. 4, pp. 147-175.

Patell, J.M./Wolfson, M.A. (1984): The Intraday Speed of Adjustment of Stock Prices to Earnings and Dividend Announcements, Journal of Financial Economics, Vol. 13, pp. 223-252.

Pettit, R.R. (1972): Dividends Announcements, Security Performance and Capital Market Efficiency, Journal of Finance, Vol. 27, pp. 993-1007.

Pettit, R.R. (1976): The Impact of Dividend and Earnings Announcements, Journal of Business, Vol. 49, pp. 86-96.

Riley, J.G. (1979): Informational Equilibrium, Econometrica, Vol. 47, pp. 331-360.

Ross, S.A. (1977): The Determination of Financial Structure: The Incentive Signalling Approach, Bell Journal of Economics, Vol. 8, pp. 23-40.

Ross, S.A. (1978): Some Notes on Financial Incentive-Signalling Models, Activity Choice and Risk Preferences, Journal of Finance, Vol. 33, pp. 777-794.

Sahling, C. (1981): Die Reaktion des Aktienmarktes auf wesentliche Ausschüttungsänderungen, Schwarzenbek.

Smith, C.W./Warner, J.B. (1979): On Financial Contracting; An Analysis of Bond Covenants, Journal of Financial Economics, Vol. 7, pp. 117-161.

Spence, A.M. (1973): Job Market Signalling, Quarterly Journal of Economics, Vol. 87, pp. 355-374.

Spence, A.M. (1974): Competitive and Optimal Responses to Signals; Analysis of Efficiency and Distribution, Journal of Economic Theory, Vol. 7, pp. 296-332.

Stiglitz, J.E. (1974): On the Irrelevance of Financial Policy, American Economic Review, Vol. 64, pp. 851-866.

Trueman, B. (1986): The Relationship between the Level of Capital Expenditures and Firm Value, in: Journal of Financial and Quantitative Analysis, Vol. 21, pp. 115-129.

Watts, R. (1973): The Information Content of Dividends, Journal of Business, Vol. 46, pp. 191-211.

Why Leasing? An Introduction to Comparative Contractual Analysis

Jan P. Krahnen and Georg Meran

Summary: The major objective of this paper is a presentation of one possible way to compare alternative contractual forms. Our methodology is applied to a prominent and still unresolved issue in finance: The explanation of the leasing contract. In contrast to previous contributions the choice set we consider is complete in a well-defined sense. It comprises rent, lease and credit arrangement. The choice between optional and fixed-terms contract is shown to depend critically on the assumptions concerning liability regime, tax system, and information distribution. Their joint consideration allows to prove the hypothesis that tax-shield differentials are neither necessary nor sufficient to establish leasing attractiveness. Both necessity part and sufficiency part of our main hypothesis contradict the more traditional case for leasing.

Table of Contents:

1. Introduction

Even a cursory look at any introductory textbook of finance reveals that the allocation of capital from financiers to firms, a seemingly simple economic task, is in fact organized according to a wide ranging variety of different contractual forms. So, for instance, the typical debt contract considered in financial theory is probably a rather crude approximation to "the many different kinds of debt" [Brealey/Myers (1984), chapter 22, or Swoboda (1981), pp. 20-48] observable in every day life. Given that financiers carefully select contract formats according to some optimality standard, the eventual task for an economic theory of contract is to demonstrate in what sense contractual choice is related to important economic parameters as, for instance, the tax system or the liability regime.

In this paper some important factors that may influence the desirability of certain types of contracts are considered. The arrangements

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

chosen are optimal in a rather restricted sense: Throughout we confine the analysis to a well-specified set of observable contractual alternatives. The focus on real-world arrangements has to be distinguished from the standard methodology applied in the principal-agent literature, where optimal arrangements between two parties usually turn out to be contingent arrangements [for a survey see Wolfstetter (1987) or Barnea/Haugen/Senbet (1985)].

In the sequel we will spell out a possible methodology for comparative contractual analysis, concentrating on a comparison between uncontingent (fixed-terms) and contingent (optional) contracts. Of course, we hope to get some insight into the circumstances under which contingent contracts appear relatively favorable. Arrangements considered comprise rent, debt, and leasing, the latter of which deserves some institutional background information.

Basically, leasing is one legal form to separate ownership of an asset from its use. It is usually distinguished between short term operate leases and long term finance leases [see Brealey/Myers (1984): 541-562 or Swoboda (1981): 31-34 for details].

An operate lease is a cancellable rent contract, which normally is accompanied by a full-service arrangement. As is well known, in case of asymmetric information about an asset's true quality, the emergence of a second hand market is seriously hampered [Akerlof (1970)]. A leasing company, then, effectively substitutes for the missing second hand market and thereby solves the adverse selection problem [Flath (1980)].

A finance lease, on the other hand, is more difficult to explain. Here, investment risk is with the lessee, since the contract is non-cancellable. During the life of the contract a fixed rental rate has to be paid periodically. In addition, at termination of the contract an option may be exercised. The option may stipulate a continuation of the contract with unchanged conditions. Alternatively the lessee may be either obliged (put option) or he deliberately chooses (call option) to purchase the leased asset at some prespecified price.

In this paper we will treat exclusively the put option variant of the long term finance lease, but the model is equally well suited to handle a call option. Besides its peculiar mix of fixed and optional elements there is a second, empirical reason for the intense academic interest in the economics of leasing. During the past fifteen or twenty years an ever rising share of the long-term debt market has been conquered by leasing arrangements. [Institutional information and data may be found in specialized journals. See for instance "Financial Management" (for the US) or "Finanzierung-Leasing-Factoring" (for West-Germany).] Actually, in most European countries more than 10% (up to 20% in the US) of overall equipment investment is financed via leasing arrangements. In 1985 aggregated bookvalues of leased assets in Germany alone amounted to some 25 billion dollars. In Germany 60% of all assets acquired under lease financing referred to real estate, notably supermarkets, office buildings, warehouses and entire industrial plants.

What accounts for the rising popularity of leasing contracts? The relevant theoretical literature has always compared leasing with its close neighbour, the credit arrangement. As was shown independently by Miller/Upton, Lewellen/Long/McConnell and Myers/Dill/Bautista in 1976, credit and leasing are basically equivalent, if financial flows are considered alone. "In an idealized competitive milieu, a reliable rationale for leasing attractiveness cannot reasonably be maintained" [Lewellen/Long/McConnell (1976), p. 797]. Of course, were there contract specific tax-shield differentials, which do not cancel out in a complete comparison[1], then leasing would indeed be attractive as a means to avoid taxation. Therefore, the major proposition stemming from a financial theory of leasing may be summarized as follows (the "Miller-Upton-et-al proposition"):

> On a competitive market a positive tax-shield differential due to the use of leasing instead of outright debt is both necessary and sufficient for leasing attractiveness.

In contrast, the main hypothesis to be defended in this paper is:

> On a competitive market a positive tax-shield differential will be neither necessary nor sufficient in order to establish leasing attractiveness.

Of course, we do not claim that Miller/Upton and all others made a mistake in deriving their result; we rather rely on a model with more

---

1) A comparison is labeled "complete", if the lessors financing decision and its associated costs are considered when calculating leasing rentals. See Miller/Upton (1976) for a very convincing presentation. The 'completeness'-argument is also the heart of Miller's seminal (1977) presidential address at the meeting of the American Finance Association.

restrictive assumptions. In our opinion, the new formulation is closer to the real world of financing and investment decision than the earlier models.

The crucial extension in our model concerns the mechanics of the capital market, which we do not assume to function costlessly. Transaction costs derive from an asymmetric distribution of information among market participants. So, for instance, a financier may never be entirely convinced about the true riskiness of the project realized by the investor. Similarly, most capital assets (machines) need careful handling and responsible maintenance while in use. Since leasing has just been described as a way to separate ownership from use, it is apparent that the unobservability of maintenance potentially is a serious problem for any financier who leases an asset to a firm.

Equiped with the asymmetric information-extension it will be investigated, whether the Miller-Upton-et-al proposition still holds. Of course, it does not. Two results merit mention. First, we show in section 3 that in the absence of any tax-differentials a case for leasing can be made. This result only holds, when the incentive problem caused by asymmetric information is aggravated by actual default risk. Therefore, information distribution and liability regime are crucial elements in the necessity part of our hypothesis. Second, in section 4 it is demonstrated that even if a positive tax-differential exists the leasing contract will not always be chosen. The sufficiency-part of our above proposition is even more irritating, if the incentive problem is disregarded altogether. In that event we are back in the world of Miller/Upton et. al., where we show that the leasing contract will in fact never be chosen. The divergence from their earlier result has a simple explanation. All previous investigations while comparing leasing to its right-hand neighbour, the debt contract, have disregarded altogether its left-hand neighbour, the rent contract.

There is a moral inhere: Any sensible comparative contractual analysis has to be very careful in selecting a complete set of available opportunities. Below we offer a more precise definition of contractual forms, which clarifies that the tupel rent , credit, and leasing is indeed complete relative to the significance of an optional element in the contract.

2. Modelling the Leasing Contract: Analytical Framework

In the subsequent analysis we focus on a financial leasing contract with the following features: The contract in general refers to a productive asset whose market value at the end of the period depends, inter alia, on maintenance effort provided by the user of the machine. Assumption 1: A contract is defined as a set (q,r), which specifies the exercise price q referring to the contingent obligation of the debtor to purchase the asset at the end of the period, and a fixed rental payment r. Both, q and r, are unrestricted apriori. Assumption 2: At the end of the period the market value g of the asset is a function of its intrinsic quality, p, and of maintenance level, m, provided by the investor: $g=\Phi(p,m)$. Since p is stochastic and since m is only privately observable, the financier only knows the market value without being able to infer maintenance effort. For simplicity we let $\Phi(p,m)$ be represented by the multiplicative expression pw(m), where p is a random quality index and w is a strictly concave, twice continuously differentiable function of maintenance effort, where $w(0)\geq 0$. Assumption 3: An asset is termed "robust" if $w'(m)=0$ for all levels of m; it is said to be "firm specific" if $w(m)=0$ for all levels of m; it is called "vulnerable" if $w'(m)>0$ for $m\geq 0$ and $w'(0)>>0$, that is if it is sensitive with respect to maintenance. Assumption 4: Real resources have to be spent in order to provide positive amounts of maintenance. The cost-of-maintenance function c(m) is strictly convex and twice continuously differentiable with $c(0)=0$ and $c'(0)$ is small. Assumption 5: The investor, with wealth v (which represents his income from other sources, $v\geq 0$), faces the usual risk in the output market, where θ is the uncertain cash flow. Since we are not interested in the investor's production decision we assume an investment project of given size, normalized to one. Assumption 6: We assume stochastic independence between θ and p and let their densities be represented by f_1 and f_2, respectively, where $f_1(\theta)>0$, all $\theta\in[\underline{\theta},\bar{\theta}]$ and $f_2(p)>0$, all $p\in[\underline{p},\bar{p}]$. For simplicity we let θ and p be uniformly distributed.

The major difference between borrowing, renting, and leasing is related to the likelihood of the put option being exercised. If the exercise price exceeds the terminal market value of the asset, q>wp, the financier will execute the option in order to maximize his income. Conversely, with wp>q, the financier will be better off by selling the asset himself on the secondary (used asset) market. Thus, a credit contract is characterized by a put option that will always be exercised, with the exercise price equaling the amount due at the end of the period (principal plus interest). A pure rental arrangement, in contrast, consists of a put option that will never be exercised, i.e. q is small, possibly zero, whereas the rental payment r is strictly positive. To summarize we define contracts by imposing specific bounds on the parameters of the model:

credit contract:	$C^c = C\ (q,r;\ \text{prob}(q \geq wp)=1)$
rent contract:	$C^r = C\ (q,r;\ \text{prob}(q > wp)=0)$
leasing contract:	$C^l = C\ (q,r;\ 0 < \text{prob}(q \geq wp) < 1)$

Table 1: Complete categorization of contractual alternatives.

In our formulation both variables, q and r, enter all contracts such that a meaningful analysis has to focus on the borderline between contracts. For, e.g. if $q \geq \bar{p}w$, the option is always exercised and changes of r and q are one-to-one. In fact, in our single period model q and r are indistinguishable from one another.

Two stochastic variables, θ in the output market and p in the input-resale market, provide sufficient structure for a unified treatment of rent, lease, and debt. Default of the investor may occur in two situations. First, generated cash flow minus costs do not cover the contractual payments, q and r, while the exercise of the option has left him with an additional "opportunity loss": pw-q<0. Second, the investor defaults on his obligations although the option has not been exercised. The realizations of θ and p determine jointly the income flows going to investor and financier, y and π, respectively. Table 2 summarizes the distribution of earnings in different states of the world, where $\hat{\theta}$ and $\tilde{\theta}$ represent bankruptcy points in case the option is, respectively is not exercised. They are defined as

(2.1) $$\hat{\theta}(p) = r + c + q - pw - v$$

(2.2) $$\tilde{\theta} = r + c - v$$

(2.3) $$\hat{p} = q/w,$$

where $\hat{p}$ is the "indifference" quality index, which makes the choice between exercising and not exercising the option irrelevant.

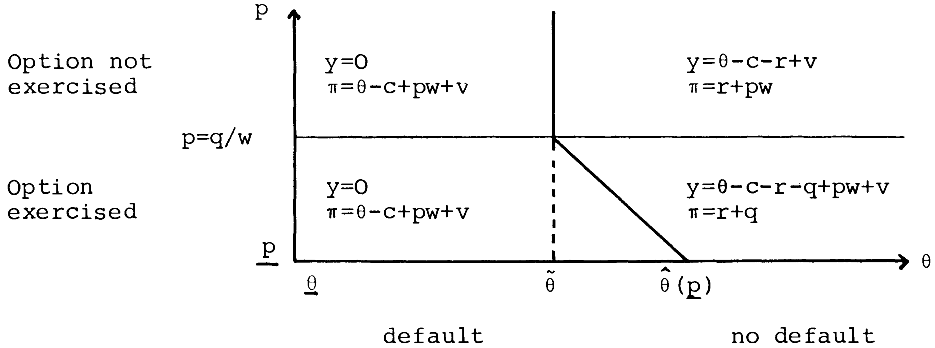


Table 2: Complete categorization of income distribution

We will assume throughout that both parties, the investor and the financier, are risk-neutral. Hence, Pareto-suboptimal risk sharing as an incentive device, which figures prominently in the analysis of the principal-agent problem (Harris/Raviv (1978), Wolfson (1985)), does not play any role in our model. For later reference we conclude this section by stating expected income of both, the investor and the financier.

$$Ey = \int_{\underline{p}}^{\hat{p}} f_2 \int_{\hat{\theta}}^{\bar{\theta}} (\theta-r-c-q+pw+v) f_1 d\theta dp + \int_{\hat{p}}^{\bar{p}} f_2 \int_{\tilde{\theta}}^{\bar{\theta}} (\theta-r-c+v) f_1 d\theta dp \qquad (2.4)$$

$$E\pi = \int_{\underline{p}}^{\hat{p}} f_2 [\int_{\underline{\theta}}^{\hat{\theta}} (\theta-c+pw+v) f_1 d\theta + \int_{\hat{\theta}}^{\bar{\theta}} (q+r) f_1 d\theta] dp + \int_{\hat{p}}^{\bar{p}} f_2 [\int_{\underline{\theta}}^{\tilde{\theta}} (\theta-c+pw+v) f_1 d\theta + \int_{\tilde{\theta}}^{\bar{\theta}} (r+pw) f_1 d\theta] dp \qquad (2.5)$$

3. The Optimal Contract with Asymmetric Information

We now will analyze the optimal contract a financier (i.e. lessor, renter or creditor) and an investor (i.e. lessee, tenant or debtor) will agree upon under the assumption of less than perfect information with respect to the decision of the investor. Specifically we assume that the investor's effort to maintain the use-value of the productive asset is unobservable to the financier. He only observes pw, its market value at the end of the period, without being able to infer from this m^*, the maintenance level privately chosen by the investor (cf. Assumption A2). We will first describe how the investor chooses his optimal maintenance level, given the contract C(q,r). In a second step the choice of the optimal contract is derived, given the reaction function of the agent. For any contract C(q,r) the investor will choose m^* by balancing returns (embodied in the resale value of the asset, pw(m)) and costs of providing that level of m, c(m). The financier, in turn, anticipates the investor's choice of m. Finding an optimal contract under asymmetric information, therefore has to incorporate an incentive compatibility requirement (ICR) describing the investor's optimal reaction given all feasible contracts. The specific ICR is found by maximizing (2.4) over m as a function of the contract. Inspection of equation (2.4) reveals that Ey is not necessarily concave in m for all values of (q,r). Our analysis, therefore, is limited throughout the paper on local effects of marginal changes of contracts. The Kuhn-Tucker conditions are (see Appendix (i) for details)

$$\frac{\partial Ey}{\partial m} = \int_{\underline{p}}^{\hat{p}} f_2 \int_{\hat{\theta}}^{\bar{\theta}} (pw'-c') f_1 d\theta dp - \int_{\hat{p}}^{\bar{p}} f_2 \int_{\tilde{\theta}}^{\bar{\theta}} c' f_1 d\theta dp \leqq 0 \tag{3.1}$$

$$\frac{\partial Ey}{\partial m} m^* = 0 \tag{3.2}$$

A sufficient condition for an interior local maximum of Ey is

$$\frac{\partial^2 Ey}{\partial m^2} = -f_2(\hat{p})\frac{qw'}{w^2}\left(\int_{\hat{\theta}}^{\bar{\theta}} \hat{p}w' f_1 d\theta\right) + \int_{\underline{p}}^{\hat{p}} (pw'-c')^2 f_1(\hat{\theta}) f_2 dp + f_1(\tilde{\theta}) \int_{\hat{p}}^{\bar{p}} c'^2 f_2 dp +$$

$$\int_{\underline{p}}^{\hat{p}} f_2 \int_{\hat{\theta}}^{\bar{\theta}} (pw''-c'') f_1 d\theta dp - \int_{\hat{p}}^{\bar{p}} f_2 \int_{\tilde{\theta}}^{\bar{\theta}} c'' f_1 d\theta dp < 0 \tag{3.3}$$

In order to derive comparative statics results of (3.1) we assume an interior solution.

$$\operatorname{sgn}\left[\frac{\partial m^*}{\partial q}\right] = \operatorname{sgn}\left[\frac{f_2(\hat{p})}{w} \int_{\hat{\theta}}^{\bar{\theta}} (\hat{p}w') f_1 d\theta - \int_{\underline{p}}^{\hat{p}} (pw'-c') f_1(\hat{\theta}) f_2 dp\right] \tag{3.4}$$

$$\operatorname{sgn}\left[\frac{\partial m^*}{\partial r}\right] = \operatorname{sgn}\left[\int_{\hat{p}}^{\bar{p}} c' f_1(\tilde{\theta}) f_2 dp - \int_{\underline{p}}^{\hat{p}} (pw'-c') f_1(\hat{\theta}) f_2 dp\right] \tag{3.5}$$

Both comparative statics results are of indeterminate sign, comprising a positive first and a negative second term. An interpretation is straightforward. In (3.4) the exercise price q is due only if pw<q. Therefore, the corresponding gains and losses to the financier caused

by a marginal increase in q center on that subset of all (θ,p) realizations, where the option is exercised. The first term in (3.4) represents the rise of the investor's expected (own) appropriation of gains due to its maintenance-investment. On the other hand, an increase of bankruptcy point $\hat{\theta}$ diminishes his expected income, since some former no-default states now have become obsolete. This explains the sign of the second term. A slightly different interpretation can be given to (3.5). The fixed rental rate r is due independent of whether the option is exercised or not. Hence, its marginal impact on the choice of maintenance centers on both bankruptcy points, $\hat{\theta}$ and $\tilde{\theta}$. The first (second) term in (3.5) quantifies the loss (the gain) caused by a marginal change of the bankruptcy points if the option is (is not) exercised.

Before we determine the optimal contract some conclusion can be drawn from the direct observation of (3.1) and (3.3).

(i) A financial arrangement C(q,r) may exist if and only if the probability of bankruptcy is less than 1, as is clear from (3.1). Note that this excludes the investor going bankrupt with certainty, when the option is not exercised, provided a contract exists.

(ii) If the probability of default equals one, in case the option is exercised, then no maintenance will be forthcoming, no matter how likely bankruptcy is, subject to the option not being exercised. This follows from the definition of $\hat{\theta}$, $\tilde{\theta}$ and equation (3.1).

(iii) If the option is always exercised and the probability of bankruptcy is zero, then the investor will choose a (first-best) Pareto-efficient effort level. To see this, insert the condition $\hat{p} \geq \bar{p}$ in equation (3.1) and obtain $Epw' - Ec' = 0$.

We are now prepared to set up the complete model. An optimal contract is a value of q and r that maximize the objective of the investor subject to the financier receiving the competitive remuneration for the capital provided, $w_o(1+i)$.

The program reads

$$\max_{q,r} Ey(m^*(q,r),q,r) \tag{3.6}$$

$$\text{s.t. } E\pi(m^*(q,r),q,r) \geq (1+i)w_o \tag{3.7}$$

$$\underline{p} \leq \hat{p} \leq \bar{p}, \tag{3.8}$$

where Ey, $E\pi$ and $\hat{p}$ are defined in (2,4), (2.5) and (2.3) respectively.

The solution to this program is spelled out in detail in our companion paper (1986). However, the major propositions and their proofs are much easier to understand if the program is transformed as follows.

Transformation

Define the function

$$q = q(r), \tag{3.9}$$

which satisfies (3.7) with strict equality. Further, insert q(r) into (2.3). This leads to

$$\hat{p} = q(r)/w(m^*(q(r),r)) \tag{3.10}$$

which determines a function

$$r = \psi(\hat{p}). \tag{3.11}$$

Utilizing (3.9) and (3.11) we are now prepared to rewrite program (3.6) - (3.8) as:

$$\max Ey(m^*(q(\psi(\hat{p})),\psi(\hat{p})),q\psi(\hat{p})),\psi(\hat{p})) \tag{3.12}$$

$$\underline{p} \leq \hat{p} \leq \bar{p}$$

The transformed program is didactically advantageous, because now all propositions can be depicted in a two-dimensional diagram, relating $\hat{p}$, the "indifferent" quality index, and Ey, the expected income of the investor. By construction the level of Ey is the feasible expected income of the contract considered.

Before we proceed in stating the main propositions it is useful to give an explanation why $\hat{p}$ is restricted to the interval $[\underline{p},\bar{p}]$.

Lemma 3.1: Let I_1 be the set of all (q,r) such that $\hat{p} < \underline{p}$, formally: $I_1 = \{q,r|\hat{p} < \underline{p}\}$. Similar define $I_3 = \{q,r|\hat{p} > \bar{p}\}$.
For all $(q,r)\in I_1$ and all $(q,r)\in I_3$ it follows:

$$dEy(\cdot)/d\hat{p} = 0.$$

Proof: See Appendix.

Lemma 3.1 is illustrated in the following diagram.

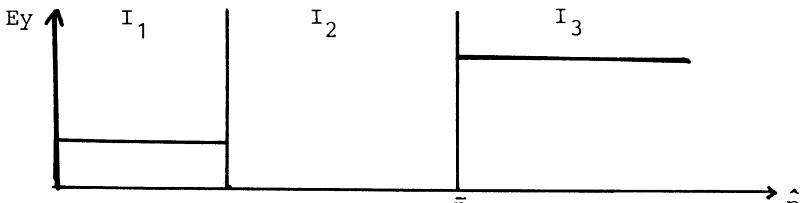


Figure 1

Fig. 1 simply shows that whenever the contract is set, such that any change of q and r has no effect on the probability of the option being exercised, then the direct incentive effect disappears (compare (3.1), (3.4) and (3.5)). Next consider the indirect effect of contractual variables (q,r) on maintenance which is due to the inclusion of default risk. Noting that (3.7) constrains all admissible pairs (q,r), it is straightforward to show that any indirect effect vanishes throughout regions I_1 and I_3.

Next we turn to a robust productive asset, whose financing is considered.

Proposition 3.1: For any robust productive asset (maintenance has no impact on the asset's physical condition) there is no unique optimal contractual form.

Proof: See Appendix.

This result can be depicted in the following diagram.

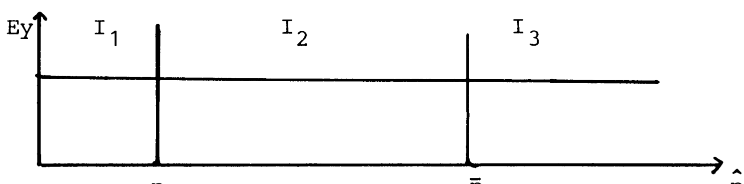


Figure 2

Whenever incentive problems with respect to maintenance do not occur, then the choice of the optimal contract is not uniquc. Rent, lease or debt-financed purchases all yield the same expected income. Of course, this result resembles the famous Modigliani/Miller (1958)-irrelevancy proposition. Note, however, that our proposition is true even if default may occur.

The remainder of this section refers to situations where an incentive problem is in fact present. To analyse the joint effect of moral hazard and default risk we first consider the no-bankruptcy case (Proposition 3.2) before turning to a complete description of the problem (Proposition 3.3).

Proposition 3.2: Suppose the market value of the asset depends on maintenance. In the absence of default risk that is, with income v sufficiently high, the optimal contract is a credit contract.

Proof: See Appendix

Figure 3 summarizes this result.

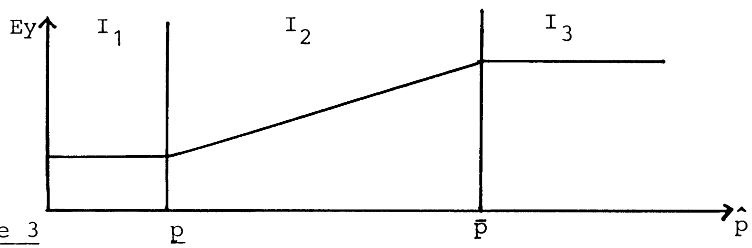


Figure 3

Observe, that the graph in I_2 is monotonically increasing which indicates that all feasible leasing contracts can be ranked according to the probability of the put option being exercised. Of course, the credit contract is optimal ($\hat{p}^* = \bar{p}$). Hence C^* has to be a credit contract. The assumed absence of default risk leaves the financier uninterested in the maintenance effort actually forthcoming. The return on his investment is a sure thing. In contrast, the investor absorbs all benefits derived from a higher maintenance level by simply owning the asset. In fact, with both actors risk-neutral and no default risk a first best contract is possible implying no risk-sharing at all (Ross (1973), Harris/Raviv (1978)).

We next turn to the joint consideration of incentive risk and default risk.

Proposition 3.3: Assume default is possible and the physical asset is sensitive with respect to maintenance. Then the optimal arrangement
(i) will never be a rent contract and
(ii) will be a leasing contract if $\partial\hat{p}/\partial q < 0$.

Proof: The proof is based on a local analysis at $\hat{p} = \underline{p}$ and $\hat{p} = \bar{p}$, respectively. See Appendix.

Once again, a figure provides an intuitive explanation.

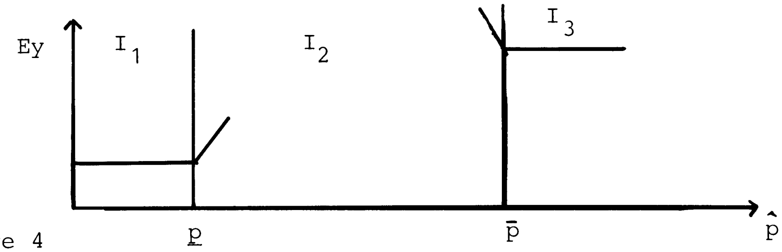


Figure 4

Case (i) is depicted on the left side of figure 4. Its interpretation is evident. Note that this result is true irrespective of whether or not default is possible. It only hinges upon the technological properties introduced in Assumptions 3 and 4.
Case (ii), however, requires some further comment. Suppose, at some value of exercise price q that an increase of q simultaneously increases maintenance (and hence resale value) and lowers the critical value $\hat{p}$. From (2.3) it follows:

$$\frac{d\hat{p}}{dq} = \frac{w - qw' \partial m^*/\partial q}{w^2} < 0 \qquad (3.13)$$

In this case, moving from $\hat{p} = \bar{p}$ to $\hat{p} < \bar{p}$ is welfare improving and hence will be chosen. But $\hat{p} < \bar{p}$ is equivalent to choosing a leasing arrangement (cf. Table 1).

4. The Impact of Taxation

Having dealt with uncertainty and incentive problems in section 3 we will now add taxation as a third relevant parameter that may influence contractual choice. Quite generally, taxation expresses the fact that the state is a silent, though ubiquitous partner of all (formal) economic activity. His claim may be codified as a fixed or a variable function of the firm's cash flow. Taxation of personal income or corporate profits, for instance, is a stepwise increasing function of realized income or profits, whereas investment tax credit in general is a function of book values, not of cash flow. In the theory of contractual choice, however, the incidence of taxation is relevant per se only to the extent that it drives a wedge between different actors. Differential taxation refers to a situation where tax burden depends inter alia on who is taxed in a given situation. Personal differences in tax burden potentially open the possibility for tax arbitrage [see the stimulating paper by Stiglitz (1983) on strategies to avoid capital gains taxation]. We will sketch the tax arbitrage argument and its impact on contractual choice, leaving the formal argument to our companion paper.

We entirely disregard uniform taxation because, quite generally, it is of little interest for partial equilibrium analysis in finance. Its introduction would not alter the relative merits of those contractual forms analysed in the last section. We hasten to add that from an economic (general equilibrium) point of view any form of taxation, be it uniform or differential, will clearly influence the overall attractiveness of investment projects and, hence, will influence the optimal investment level. Since here we do not consider investment behavior, but in fact treat cash flow from regular business, θ, as given, it follows that the analysis of contractual choice is not affected by any form of uniform taxation. In the sequel we will briefly review the existing literature on the subject and then go on to show where the traditional case for leasing, emanating from tax considerations, is basically incomplete and indeed misleading.

In line with traditional financial economics of the fifties and early sixties, a first generation of authors conducted a partial analysis of leasing arrangements. By exclusively looking at the investor's choice between lease or purchase it was straightforward to demonstrate leasing attractiveness by pointing at the seemingly "cheaper" lease. It was only in the mid-seventies that a more complete modelling of the lease versus buy decision emerged. This second generation jointly considered investment <u>and</u> financing. It became immediately apparent that the "cheap" lease disappears when its rental rates are endogenously determined, thereby including capital costs of the financier. The irrelevance with respect to contractual choice is lost again, however, if the firm using a certain asset and the firm financing it are subject to different tax rates (see Myers/Dill/Bautista (1976) for a complete derivation).

As already mentioned, differential taxation is commonly considered the single most important parameter explaining the success of leasing. Since in finance literature incentive problems are usually not considered, we will trace the impact of taxation first for the case of symmetric information. The risk of default, however, will be included in this first stage. Arguments in favor of leasing usually rely on taxation of invested capital ("investment tax credit" in the US, "Gewerbekapitalsteuer" in West Germany). In addition to assumption 3.1 through 3.6 (see beginning of section 2) we make

Assumption 4.1 (tax code): Taxes are levied upon the market value of the productive asset at the end of the period, pw.

This certainly is an idealization of tax reality, as assumption 4.1 implies valuation for tax purposes always to be done at (equivalent) replacement costs[1]. In addition it is assumed that taxes on invested capital represent preferred debt vis-a-vis the public, which gives it a priority claim in any bankruptcy proceeding. Since we model a one-period setting, loss-offset or carryforward provisions are not considered. Finally it is differentiated between the tax rate applying to the investor, t, and to the financier, t^*.

Table 3 is a distribution diagram, representing income of financier and investor as a function of the joint realization of quality index p and project cash flow θ.

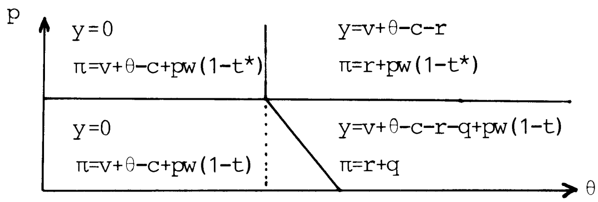


Table 3: Complete categorization of income distribution with differential taxation

Once again, the upper region (all p-θ realizations above $p=\hat{p}$) corresponds to a situation where the ownership of an asset after termination of the one period contract remains with the financier. This may be due to either ex-ante (rent contract) or ex-post choice (leasing contract). Conversely, with $p<\hat{p}$ the option will be exercised such that ownership remains with the investor, which again may be due to ex-ante choice (credit contract) or ex-post choice (leasing-contract). The tax differential implies that whenever the option is (is not) exercised, then the higher rate t (the lower rate t^*) is applicable. Note the following peculiarity of our tax code: If the

1) German fiscal practice, for instance, is to tax on the basis of rateable value ("Einheitswert") of an asset, where correspondence with replacement costs of an equivalent asset is reviewed after some interval only. A careful application of the traditional Miller-Upton-et-al-proposition to the peculiarities of the German tax code can be found in Schröder (1984).

option is exercised and the investor declares bankruptcy, then the financier, although personally in the lower tax bracket, has to assume the investor's debt vis-a-vis the treasury. Therefore, in the lower left region of the table 3 the financier is taxed at the appropriate rate t

By an analysis similar to section 3, two major propositions can be proved.

Proposition 4.1

Suppose, information is private, default is possible ($\underline{\theta} \leq \tilde{\theta} \leq \hat{\theta} \leq \bar{\theta}$), the asset is robust (w'=O) and a tax differential in favor of the financier exists ($t^* < t$). Then the optimal contract is a rent contract.

Looking at the optimal contract in the absence of incentive problems, but when default is possible, corresponds to the setting of proposition 3.1. Not surprisingly, we find that the Modigliani/Miller result no longer holds. Intuitively, differential taxation invites investors to separate economic usufruct from ownership title in order to avoid taxation at the higher rate t.

It is straightforward to show that now the Miller-Upton-et-al-proposition is valid, in that credit is dominated by a leasing arrangement. But the analysis does not stop here because a more preferred alternative, the rent contract, is available.

Proposition 4.1 directs attention to an assumption implicit in prevailing proofs of a tax-rationale for leasing: It has always been supposed that the lessee will in fact never been taxed at the higher corporate rate t, not even after termination of the contract and, hence, after the eventual acquisition of the asset.

We next turn to a situation where the incentive problem is considered.

Proposition 4.2

Assume information is private, the asset is vulnerable (sensitive with respect to maintenance), default is excluded (v>>O) and a positive tax differential exists ($t^* < t$). Then the optimal contract may be of the rent, leasing or credit variety depending on (i) the size of the tax-differential and (ii) the size of ϵ, the elasticity of the asset's resale value with respect to exercise price q[1]. The assumptions correspond to those of proposition 3.2, where the global optimality of the credit contract was established. The case for credit is now considerably weakened. As a matter of fact, credit will now be chosen if and only if $\epsilon=1$. Whenever $\epsilon<1$, the determination of an optimal contract is related to the size of the tax differential. For $[t-t^*]$ large, the incentive effect of a contingent contract is outweighted by the permanent tax advantage of the uncontingent rent contract. On the other hand, low tax differential favor contracts which bears on maintenance incentives, i.e. a leasing arrangement.

Both propositions, 4.1 and 4.2, emphasize the sufficiency-part of our main hypothesis, presented in the introduction: although an exploitable tax advantage is assumed, a leasing arrangement need not be the optimal arrangement.

5. Discussion

This paper introduces a method to compare optional and fixed contracts. In contrast to earlier literature analysing the leasing decision, we (i) explicitly consider the optional element in the long term financial lease, and (ii) describe a complete set of alternatives, comprising rent, lease and debt contracts. These arrangements are compared with respect to their handling of three crucial conditions determining contractual choice: Information structure (incentive risk), liability rule (default risk), and tax code (differential taxation). It has been established that a rationale for leasing attractiveness cannot be maintained if any one crucial condition is considered alone. However, a joint consideration of informational asymmetry and either default risk (section 3) or tax differential (section 4) proves to be sufficient.

Of course, we are aware that our analysis has not produced a convincing explanation of the rising attractiveness of leasing arrangements on real-world capital markets. However, we have tried hard and, in fact, experimented with a variety of different model specifications,

1) In order to derive ϵ, recall that $\hat{p}\equiv q/w(m(q,r))(1-t^*)$. Therefore, $d\hat{p}/dq = [w/(1-t^*)-qw'(dm^*/dq)(1-t^*)]/[w^2(1-t^*)^2]=[1/w(1-t^*)][1-\epsilon]$, where $\epsilon\equiv(q/w)(dw/dq)$, and $d\hat{p}/dq \gtrless 0$ as $\epsilon \lessgtr 1$.

but a strong case for leasing has not been spotted.

Possibly, we may have asked the wrong question. Recent empirical evidence points in that direction. Analysing the leasing decision of some 600 US-corporations, Ang/Peterson (1984) find that leasing and debt are complements for one another rather than substitutes, as our model has assumed. One may therefore attempt to model the lease decision in the context of capital structure optimization. A rationale for leasing that corresponds to observed complementarity may once again concentrate on its distinguishing contractual feature: the conditional separation of ownership and use.

The story could run as follows.
Consider an informationally imperfect financial market. It is well known that here lenders will wish to ration their customers [see Baltensperger/Devinney (1985) for a survey]. Effective rationing, however, requires the bank to be perfectly informed about a customer's remaining access to the financial market . Now assume that information about each customer's additional sources of debt (his "liquidity reserve") is imperfect as well. Then, of course, rationing is not feasible any more, because the borrower's ability to substitute between different lenders makes any single threat of debt cancellation incredible.
In a like situation a leasing contract may be valuable. Note that a cancelled lease cannot immediately be substituted on the credit market, because it implies a physical disinvestment of the asset itself. Real production will be interrupted until a substitute machine is installed, thereby affecting the firm's cash flow.

Of course, if this story were true, we would expect lease financing to account for a certain fraction of total corporate debt. Complementarity of debt and leasing in corporate capital structure, then, is required to render the threat credible.

6. Appendix

(i) Maximizing Ey with respect to m:

First rewrite (2.4) as follows:

(A.1) $$\int_{\underline{p}}^{\hat{p}} \hat{B}(\hat{\theta},p,m)f_2 dp + \int_{\hat{p}}^{\bar{p}} \tilde{B}(\tilde{\theta},m)f_2 dp$$

when

(A.2) $$\hat{B}(\hat{\theta},p,m): = \int_{\hat{\theta}}^{\bar{\theta}} (\theta-r-c-q+pw+v)f_1 d\theta$$

and

(A.3) $$\tilde{B}(\tilde{\theta},m): = \int_{\tilde{\theta}}^{\bar{\theta}} (\theta-r-c+v)f_1 d\theta$$

Note, that c, w, $\hat{\theta}$ and $\tilde{\theta}$ are functions of m. Now, utilizing the Leibnitz-rule, the differentiation of (A.1) yields:

(A.4) $$\frac{-qw'}{w^2} f_2(\hat{p})\hat{B}(\hat{\theta},\hat{p},m) + \int_{\underline{p}}^{\hat{p}} \left(\frac{\partial \hat{B}}{\partial \hat{\theta}} \frac{\partial \hat{\theta}}{\partial m} + \frac{\partial \hat{B}}{\partial m}\right) f_2 dp + \frac{qw'}{w^2} f_2(\hat{p})\tilde{B}(\tilde{\theta},m) + \int_{\hat{p}}^{\bar{p}} \frac{\partial \tilde{B}}{\partial m} f_2 dp$$

Recalling the definitions

(2.1) $$\hat{\theta}(p) = r+c(m)+q-pw(m)-v$$

(2.2) $$\tilde{\theta} = r+c(m)-v$$

(2.3) $$\hat{p} = q/w(m)$$

it follows by (A.2) and (A.3), that

$$\hat{B}(\hat{\theta},\hat{p},m) = \tilde{B}(\tilde{\theta},m)$$

and

$$\partial\hat{B}/\partial\hat{\theta} = 0 .$$

Hence (A.4) reduces to

(A.5) $$\int_{\underline{p}}^{\hat{p}} \frac{\partial \hat{B}}{\partial m} f_2 dp + \int_{\hat{p}}^{\bar{p}} \frac{\partial \tilde{B}}{\partial m} f_2 dp$$

where, by (A.2) and (A.3),

$$\frac{\partial \hat{B}}{\partial m} = \int_{\hat{\theta}}^{\bar{\theta}} (pw'-c')f_1 d\theta$$

and $$\frac{\partial \tilde{B}}{\partial m} = \int_{\tilde{\theta}}^{\bar{\theta}} c'f_1 d\theta.$$

(A.5) is the left hand side of (3.1). The corresponding second order condition (equation (3.3)) is obtained by applying the Leibnitz-rule to (A.5).

(ii) Solving program (3.12):

In order to work out the optimality conditions for program (3.12) it is necessary to differentiate functions

(3.9) $q=q(r)$ with respect to r

and

(3.11) $r=\psi(\hat{p})$ with respect to $\hat{p}$.

q' can be obtained by implicitly differentiating the capital market constraint (3.7):

$$\text{(A.6)} \qquad q' = -\frac{E\pi_m m_r^* + E\pi_r}{E\pi_m m_q^* + E\pi_q}$$

with

$$\text{(A.7)} \qquad E\pi_m = \int_{\underline{p}}^{\hat{p}} f_2 \int_{\underline{\theta}}^{\hat{\theta}} (pw'-c')f_1 d\theta dp + \int_{\hat{p}}^{\bar{p}} f_2 [\int_{\underline{\theta}}^{\tilde{\theta}} (pw'-c')f_1 d\theta + \int_{\tilde{\theta}}^{\bar{\theta}} pw' f_1 d\theta]$$

$$\text{(A.8)} \qquad E\pi_q = \int_{\underline{p}}^{\hat{p}} f_2 \int_{\hat{\theta}}^{\bar{\theta}} f_1 d\theta dp$$

$$\text{(A.9)} \qquad E\pi_r = \int_{\underline{p}}^{\hat{p}} f_2 \int_{\hat{\theta}}^{\bar{\theta}} f_1 d\theta dp + \int_{\hat{p}}^{\bar{p}} f_2 \int_{\tilde{\theta}}^{\bar{\theta}} f_1 d\theta dp$$

Similar, ψ' can be obtained from (3.10) by implicit differentiation:

$$\text{(A.10)} \qquad \psi' = \frac{1}{(\partial\hat{p}/\partial q)q' + \partial\hat{p}/\partial r},$$

where

$$\text{(A.11)} \qquad \partial\hat{p}/\partial q = \frac{1}{w^2}[w - qw' m_q^*]$$

$$\text{(A.12)} \qquad \partial\hat{p}/\partial r = \frac{-1}{w^2}[w' q m_r^*]$$

To solve program (3.12) we define the Lagrange-function

$$\text{(A.13)} \qquad L = E\pi + \sigma_1[\hat{p}-\underline{p}] + \sigma_2[\bar{p}-\hat{p}]$$

and maximize it with respect to $\hat{p}$, (minimize it with respect to σ_1, σ_2).

This yields the following Kuhn-Tucker-conditions:

(A.14) $\hat{p}$: $\psi'[Ey_q q' + Ey_r] - \sigma_2 + \sigma_1 = 0$

(A.15) σ_1: $[\hat{p} - \underline{p}] \geq 0$ and $[\hat{p} - \underline{p}]\sigma_1 = 0$

(A.16) σ_2: $[\bar{p} - \hat{p}] \geq 0$ and $[\bar{p} - \hat{p}]\sigma_2 = 0$

where

(A.17) $$Ey_q = - \int_{\underline{p}}^{\hat{p}} f_2 \int_{\hat{\theta}}^{\bar{\theta}} f_1 d\theta dp$$

(A.18) $$Ey_r = - \int_{\underline{p}}^{\hat{p}} f_2 \int_{\hat{\theta}}^{\bar{\theta}} f_1 d\theta dp - \int_{\hat{p}}^{\bar{p}} f_2 \int_{\tilde{\theta}}^{\bar{\theta}} f_1 d\theta dp$$

From (A.8) and (A.9) we have

(A.19) $Ey_q = -E\pi_q$

(A.20) $Ey_r = -E\pi_r$

<u>(iii) Three preliminary results:</u>

<u>Lemma A.1</u>

If a rent contract is adopted, then by assumption A.3 und A.4,

$$\partial\hat{p}/\partial q \Big|_{\hat{p}=\underline{p}} > 0.$$

<u>Proof:</u> From (2.3) it follows

$$\partial\hat{p}/\partial p \Big|_{\hat{p}=\underline{p}} = [w - qw'(\partial m^*/\partial q)]/w^2$$

which by (3.3) and (3.4) is positive provided that c'(O) is small.//

Lemma A.2

If bankruptcy cannot be excluded ($\underline{\theta} < \tilde{\theta} \leq \hat{\theta}(p)$), then

(a) $$\int_{\underline{p}}^{\bar{p}} f_2(p)(pw'-c')f_1(\hat{\theta})dp > 0$$

and

(b) $$\int_{\underline{p}}^{\bar{p}} f_2(p) \int_{\underline{\theta}}^{\hat{\theta}} (pw'-c')f_1(\theta)d\theta dp > 0,$$

where $\hat{\theta}$ is defined in (2.1).

Proof:

Define (i) $A(p) \equiv f_2(p)(pw'-c')$

(ii) $\hat{f}_1 \equiv f_1(\hat{\theta})$

(iii) $\hat{F}_1 \equiv \int_{\underline{\theta}}^{\hat{\theta}} f_1(\theta)d\theta$

Recall from (2.1) that $\hat{\theta}$ is a function of p. Rewriting the first term of (3.1) using (i) to (iii) and expanding a Taylor series around p^*, where p^* is chosen such that $A(p^*) = 0$. we get

(iv) $$\int_{\underline{p}}^{\bar{p}} A(p)(1-\hat{F}_1)dp = w'[(1-\hat{F}_1)\int_{\underline{p}}^{\bar{p}}(p-p^*)dp - \hat{f}_1 w\int_{\underline{p}}^{\bar{p}}(p-p^*)^2dp] = 0$$

Similarly, a Taylor expansion of (a) und (b) around p^* yields

(v) $$\int_{\underline{p}}^{\bar{p}} A(p)\hat{f}_1 dp = w'\hat{f}_1\int_{\underline{p}}^{\bar{p}}(p-p^*)dp$$

(vi) $$\int_{\underline{p}}^{\bar{p}} A(p)\hat{F}_1 dp = w'[\hat{F}_1\int_{\underline{p}}^{\bar{p}}(p-p^*)dp + \hat{f}_1 w\int_{\underline{p}}^{\bar{p}}(p-p^*)^2dp]$$

By (iv), both (v) and (vi) are positive.//

Lemma A.3

If bankruptcy never occurs, $\partial\hat{p}/\partial q > 0$.

Proof: From (A.11) we know that

$$\partial\hat{p}/\partial q = [w-qw'(\partial m^*/\partial q)]/w^2,$$

which, by (3.3) and (3.4) is positive.//

(iv) Proof of results in section 3:

Proof of Lemma 3.1:

Utilizing (3.4) and (3.5) and Lemma A.2 we easily can find

$$m_q^* = m_r^* \geq 0, \ \forall (q,r) \in I_1 \cup I_3. \qquad \text{(i)}$$

Insert this result, (A.6) - (A.12) and (A.17) - (A.20) into the first order condition (A.14) and delet both σ_1 and σ_2. Then one easily can observe that (A.14) vanishes $\forall (q,r) \in I_1 \cup I_3$, which proves the assertion.//

Proof of Proposition 3.1:

Since $w' = 0$, it follows by (3.2) that no maintenance effort is forthcoming, i.e. $m^* = 0$, $\forall (q,r)$ and hence $m_q^* = m_r^* = 0$. Inserting this result into (A.6) and (A.10) - (A.12) yields the first order condition (A.9) which reduces to

$$w[Ey_q \frac{E\pi_r}{E\pi_q} + Ey_r] - \sigma_2 + \sigma_1 = 0, \qquad \text{(i)}$$

where, by (A.19) and (A.20), the term in bracket vanishes, $\forall \hat{p}$, and hence by (A.15) and (A.16) $\sigma_2 = \sigma_1 = 0$. This proves that neither an interior nor a corner solution is uniquely optimal.//

Proof of Proposition 3.2:

We first show that the first order conditions (A.14) - (A.16) are satisfied if and only if $\hat{p} = \bar{p}$, i.e. C^* is a credit contract. First evaluate Ey_q and Ey_r from (A.17) and (A.18) for $\tilde{\theta} \leq \hat{\theta} < \underline{\theta}$. Inserting these terms into (A.14) yields

$$\psi'[(q'-1)(\int_{\underline{p}}^{\hat{p}} f_2 dp) - \int_{\underline{p}}^{\bar{p}} f_2 dp] = \sigma_2 \qquad \text{(i)}$$

with $\sigma_1 = 0$ since a rent contract will never be optimal as it will be shown in Proposition 3.3, i.

Utilizing (A.6) - (A.10) (i) can be written as

$$\psi'[\frac{\int_{\underline{p}}^{\hat{p}} f_2 dp}{\int_{\hat{p}}^{\bar{p}} f_2 dp m_q^* + \int_{\underline{p}}^{\hat{p}} f_2 dp} - 1] = \sigma_2, \qquad \text{(ii)}$$

where ψ' reduces to

$$\psi' = 1/[\partial\hat{p}/\partial q)q'] < 0. \qquad \text{(iii)}$$

Note that the term in bracket of eq. (ii) is negative(zero) for $\hat{p} < \bar{p}$ $(\hat{p}=\bar{p})$.
Now, assume, by contradiction, that $\hat{p}^* < \bar{p}$. Then from (ii) and (iii) it follows

$$\sigma_2 > 0,$$

which contradicts the optimality condition (A.16). Therefore $\hat{p}^* = \bar{p}$, i.e. the optimal contract C^* must be a credit contract. To prove that the second order condition is satisfied at $\hat{p} = \bar{p}$, differentiate (ii) with respect to $\hat{p}$. This leads to

$$\psi' f_2(\hat{p}) m_q^*,$$

which, by (iii) and (3.4) is negativ.//

Proof of Proposition 3.3.(i):

Assume, by contradiction, that the optimal contract C^* were a rent contract, i.e. $\hat{p}^* = \underline{p}$. Hence, by (3.1) $m^* = 0$. Note, however, by the assumed properties of w and c, that $m_q^* > 0$ and $m_r^* \geq 0$ at $\hat{p} = \underline{p}$. The first order condition (A.14) reduces now to

$$\psi' E y_r = - \sigma_1. \qquad \text{(i)}$$

Recalling (A.6) - (A.9) and (A.11), (A.12), (i) can be written as:

$$E y_r / [- \frac{\partial\hat{p}}{\partial q}(E\pi_m m_r^* + E\pi_r) + \frac{\partial\hat{p}}{\partial r}] = - \sigma_1 E\pi_m m_q^* \qquad \text{(ii)}$$

Note from (A.18), (A.7), (A.9) that $E\pi_r > 0$, $Ey_r < 0$ and $E\pi_m > 0$ for $w'(0) >> 0$ and $c'(0)$ small. Also, by Lemma A.1, $\partial\hat{p}/\partial q|_{\hat{p}=\underline{p}} > 0$. Hence, for $\sigma_1 \geq 0$, (ii) is a contradiction which proves that C^* cannot be a rent contract.//

Proof of Proposition 3.3. (ii):

The proof proceeds as follows. First assume, by contradiction, that C^* is a credit contract and evaluate all relevant equations at $\hat{p} = \bar{p}$. Utilizing (A.7) - (A.9) and (A.17), (A.18) it is then an easy task to rewrite the first order condition (A.14) as follows:

$$- \psi' E\pi_q \alpha = \sigma_2, \qquad \text{(i)}$$

where

$$\psi' > 0 \tag{ii}$$

and

$$\alpha := (q'+1) > 0 \tag{iii}$$

Obviously, this produces a contradiction for all $\sigma_2 \geq 0$. We therefore have to prove (ii) and (iii). First, we prove (ii). From (A.10) - (A.12) we have

$$\text{sign}[\psi'] = \text{sign}[\beta] \tag{iv}$$

where

$$\beta := (w - qw' m_q^*)q' - qw' m_r^* \tag{v}$$

The assumption $\partial\hat{p}/\partial q < 0$ determines the positive sign of m_q^*. From (3.3), (3.5) and Lemma A.2 we find $m_r^* < 0$. Also, note that $q'(r)$ cannot be positive. Otherwise the capital market constraint (3.7) would not be binding since a reduction of both q and r according to q(r) is in the interest of the investor while not affecting the expected return of the financier. Hence, from (v), (A.11) and (A.12) it follows that $\beta > 0$ and therefore, by (iv), $\psi' > 0$.
To prove (iii), recall (A.6) and insert (A.7) - (A.9). Then, it immediately follows:

$$q' = -\frac{E\pi_m m_r^* + E\pi_q}{E\pi_m m_q^* + E\pi_q} > -1.$$

Hence, by (iii) $\alpha > 0$.//

References

Akerlof, G. (1970), "The market for 'lemons': Quality uncertainty and the market mechanism", Quarterly Journal of Economics, 84, 488-500.

Ang, J. and P. Peterson(1984), "The leasing puzzle", Journal of Finance, 39, 1055-1065.

Baltensperger, E. and T. Devinney (1985), "Credit rationing theory: A survey and synthesis", Zeitschrift für die gesamte Staatswissenschaft, 141, 475-502.

Barnea, A., Haugen, R. and L. Senbet (1985), Agency problems and financial contracting, Prentice Hall.

Brealey, R. and S. Myers (1984), Principles of corporate finance, McGraw Hill.

Flath, D. (1980), "The economics of short-term leasing", Economic Inquiry, 18, 247-259.

Harris, M. and A. Raviv (1979), "Optimal incentive contracts with imperfect information", Journal of Economic Theory, 20, 231-259.

Krahnen, J. and G. Meran (1986), "The buy-versus-lease-versus-rent decision: A comparative contractual analysis", Discussion Papers on Political Economy No. 37, Freie Universität Berlin.

Lewellen, W., Long, M. and J. McConnell (1976), "Asset leasing in competitive capital markets", Journal of Finance, 31, 787-796.

Miller, M. (1977), "Debt and taxes", Journal of Finance, 32, 261-275.

Miller, M. and C. Upton (1976), "Leasing, buying, and the cost of capital", Journal of Finance, 31, 761-786.

Myers, S., Dill, D. and A. Bautista (1976), "Valuation of financial lease contracts", Journal of Finance, 31, 799-819.

Ross,S. A.(1973), "The economic theory of agency: The principal's problem", American Economic Review, 63, 134-139.

Schröder, J. (1984), Die Stichhaltigkeit von Argumenten für und wider Leasing, Harri Deutsch.

Stiglitz,J.E.(1983), "Some aspects of the taxation of capital gains", Journal of Public Economics, 21, 257-294.

Swoboda, P. (1981), Betriebliche Finanzierung, Physica.

Wolfson, M. (1985), "Tax, incentive, and risk-sharing. Issues in the allocation of property rights: The generalized lease-or-buy problem", Journal of Business, 58, 159-171.

Wolfstetter, E. (1987), "Contract theory and unvoluntary unemployment: A review", Zeitschrift für Wirtschafts- und Sozialwissenschaften, 107, forthcoming.

The Financial Theory of Agency as a Tool for an Analysis of Problems in External Accounting

Ralf Ewert

Summary: The purpose of this paper is to show how the financial theory of agency may be fruitfully applied to some problems in the field of external accounting. It is argued that a solution to certain normative problems in financial accounting requires a theory, which explicitly deals with the consequences of conflicts of interest between various groups of financiers for the investment and financing policy of the firm. The paper first shows that the financial theory of agency may provide such an approach. Then some results of existing models are surveyed, especially with regard to the interdepencies between the role of accounting systems in mitigating agency problems of debt and several financing scenarios.

Table of Contents:

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

1. Introduction

Although the investigation of problems in external accounting has a long history in economic science, it seems that there are still plenty of questions which have not yet been treated and answered in a satisfactory way. Studies about economic effects of external accounting systems or valuation principles often prove to be value judgements or opinions rather than rigorous analyses. This disenchanting situation was until the mid-seventies partly due to a lack of economic models which were able to explicitly address central problems relevant to external accounting. In observing this fact, Hakansson (1978), p. 724, stated that "advances in finance, economics, and behavioral science are in the nature of a pre-condition for substantial further progress in accounting." This paper wants to show, that the Financial Theory of Agency (FTA), which has rapidly emerged from the pathbreaking article of Jensen/Meckling (1976), may constitute such an advance in finance.

This assertion does not mean that the FTA solves all problems in accounting nor that it constitutes the universal approach all accounting researchers and practitioners had waited for. Such a claim would neither be credible nor reasonable, because the theory is yet - despite the great progress achieved during the last years - just in its first stages. But it is argued that the FTA presents a framework which enables a detailed and rigorous analysis of *some* problems of great interest to accounting researchers, especially when the function of an accounting system is viewed as the "protection of creditors' interests", which is one of the major goals of the external accounting system regulated by law in West Germany.[1]

This paper is organized as follows: Section 2 gives a basic description of the FTA with special emphasis on the debt-related agency problems. This does not mean that the agency problems of equity are irrelevant for external accounting, but with regard to the development of concrete models for integrating the FTA and external accounting, the debt-related agency problems have up to now received greater attention in literature than the agency problems of equity, so that this restraint becomes plausible. Section 3 shows how the FTA can generally be applied to positive and normative accounting questions.[2] Section 4 presents some re-

---

1) The most recent discussion of balance-sheet purposes with respect to West Germany appears in Ellerich (1986). Note, however, that balance-sheet purposes differ from country to country. For example, the "protection of creditors' interests is not a goal for the regulated accounting system in the USA.

2) With respect to applications of other branches of the agency literature in the field of accounting see, for example, Gjesdal (1981).

sults of existing models. Because of the limited scope of this article it is impossible to present the models in section 4 in full depth, so that only the basic ideas for the derivation of the results will be outlined and mathematical reasoning is kept to a minimum for the purpose of this paper.

2. The Financial Theory of Agency (FTA): A Basic Characterization

2.1 Agency Relationshipsin Finance and the Tasks of the FTA

According to Jensen/Meckling (1976), p. 308, an agency relationship can generally be defined as "a contract, under which one or more persons (the principal(s)) engage another person (the agent) to perform some service on their behalf which involves delegating some decision making authority to the agent".[1)] Although this definition covers only the case of one agent, it should be clear, that a contract with the same characteristics as mentioned above except that there are many agents constitutes an agency relationship, too. It is one of the main features of such relationshipsthat after the parties have entered into the contract and time passes by, the agents regularly do not bear the full consequences of their actions, because in most cases they share in some form or another the outcomes of their actions with the principal(s). Since it is assumed that all individuals act as only self-interested maximizers of personal welfare, the actions chosen by the agents will therefore depend upon the sharing rule and be altered compared to a situation where they bear the full consequences. These aspects have to be taken into account by the principal(s) in determining the terms of the contract.

The above definition is very general and allows for many contractual relations to be interpreted as agency relationships. In their seminal article Jensen/Meckling (1976) interpreted the relations between an owner-manager of a firm (that is a manager who holds a positive fraction of the firm's shares) and external financiers (creditors and shareholders who are not managers) as agency contracts. In this view the owner-manager ist the agent and the external financiers are the principals. Conflicts of interest between these groups emerge from the *structure of the claims* against the firm, that is, the fraction of shares held by the owner-manager and the amount of creditors' claims. This financing structure determines a special kind of *sharing rule* and so influences the decisions which are optimal from the viewpoint

---

1) A similar definition appears in Ross (1973), p. 134.

of the agent after the money has been received from the external financiers.

Viewing financing relations in this way leads to a cancellation of one of the major assumptions in deriving the irrelevance propositions in finance[1]: The outcomes of the firm are no longer independent of the financing decisions, not only because of potential tax effects, but also for reasons of conflicts of interest, which may result in another investment and financing program than otherwise (that means including potential tax effects but excluding the influence of the sharing rule itself) would be optimal. With this background, the FTA tries to accomplish three tasks:

(a) The first one is to study the influences of the sharing rule itself by a precise analysis of the agent's optimal actions if a certain capital structure is assumed and the firm has already got the money. This results in an identification of several wealth transfer mechanisms (agency problems), which the agent may use to promote his own well-being at the expense of the other financiers' welfare after he has received the money.

(b) The second task is to establish a framework, in which it is possible to identify that group which *ultimately* bears the welfare losses resulting from agency problems and which is therefore generally interested in an installment of mechanisms to mitigate these problems.

(c) The third task is to study the functioning of several instruments which could be designated to resolve agency problems and by this means to obtain a set of instruments, which is suitable for a mitigation of agency problems in a given situation.

The remainder of this chapter follows the above mentioned tasks: Firstly a description of the usually studied agency problems will be given. Secondly the framework for an analysis of the second and third task will be sketched. For a more detailed discussion of the concepts presented below, especially with regard to the agency problems of equity, the reader is referred to the reviews of Barnea/Haugen/Senbet (1981), (1985) and Jensen/Smith (1985).

2.2 A Description of Agency Problems

(a) Agency Problems of External Equity

Imagine an owner-manager who derives utility from the three sources

---

1) See Modigliani/Miller (1958); Stiglitz (1969); Schall (1972); Stiglitz (1974); Fama (1978).

(α) money wages (which henceforth are assumed to be fixed), (β) market value of the firm's shares hold by himself (γ) and nonpecuniary benefits, which are assumed to be job-specific and inseparable from the firm (the manager may, for example, value a luxurious office, the power over the firm's employees and/or the social prestige which is often interconnected with enlarging the firm's resources and thus managing a great enterprise). Assume furthermore that no debt is used. If the manager holds 100% of the firm's shares he bears the full cost of extending the amount of nonpecuniary benefits beyond the market-value-maximizing level in the form of a reduction in the value of his shares, which equals the reduction in the firm's market value. If he holds only a fraction of the equity, he derives the same utility from extending the nonpecuniary benefits, but no longer bears the full cost. Therefore his consumption of nonpecuniary benefits in the latter situation will be greater than in the former one, which corresponds to a reduction in the firm's market value.

This situation was the one originally described and further elaborated by Jensen/Meckling (1976). However the conflicts of interest between managers and external stockholders are much broader in scope and emerge from several sources, which cannot be studied in this paper (for a review see Jensen/Smith (1985), pp. 101 - 111). For reasons mentioned above, the remainder of this article concentrates on the bondholder-stockholder conflict.

(b) Agency Problems of Debt

The following description relies on the assumption that the manager acts in the stockholders' interest, which at any time is assumed to be the maximization of the with-dividend market value of the firm's shares. There is limited liability for stockholders and a situation in considered where the bondholders' funds are already raised. The goal is to identify possible wealth transfermechanisms, i. e., to indicate actions which either become profitable for stockholders only because they harm bondholders or become unprofitable for stockholders only because the benefits partly accrue to bondholders. The agency problems of debt usually studied in literature emerge from the following sources:

(b1) Asset Liquidation and Payout of the Proceeds[1)]

---

1) See Black/Scholes (1973), p. 651; Black (1976), p. 7; Myers (1977), pp. 162-163. Potential conflicts of interest between stockholders and *customers* with respect to liquidation decisions are studied by Titman (1984) and Swoboda (1987, this volume).

If there is no payout constraint, stockholders may choose to liquidate the firm's assets and to pay themselves the proceeds as a dividend. That action will be profitable for stockholders if the assets' liquidating value exceeds the value of the shares in the going concern case (note that the *total* market value of the firm as a going concern is irrelevant for this decision rule) and the creditors' claims would be worthless.

(b2) Debt-Financed Dividend Payments[1)]

Stockholders may raise funds from new bondholders granting them priority of equal or higher[2)] order than the initial debt claims and distribute these funds to themselves. If the investment policy does not change and potential signalling and tax effects are ignored, the total market value of the firm remains constant, but wealth is transferred from the old bondholders to the stockholders[3)]. The reason is, that - due to the above mentioned priority rule - the new bondholders' claims get value from partly expropriating the initial debt. Since the new creditors pay for the value they receive from their claims and because stockholders get these funds, a wealth redistribution occurs.

(b3) Asset Substitution ("Risk Incentive Problem")[4)]

This problem can best be demonstrated in a situation, where it is assumed, that investment takes place at time t and liquidation occurs in t + 1. Remember the sharing-rule-feature of the debt contract: Since the creditors' claims have priority over the stockholders' claims, the operating cash flows in t + 1 first of all accrue to bondholders and only the residual amount is left to stockholders. However, bondholders can never get more than their contracted claims, while they sometimes suffer losses if these claims exceed the operating cash flows. Therefore stockholders - after having raised funds from creditors - might find it profitable to alter the risk of the investment program by sub-

---

1) See Fama/Miller (1972), pp. 151-152; Kim/McConnell/Greenwood (1977); Krainer (1977), pp. 2-4; Smith/Warner (1979), p. 118.

2) This can be achieved, for example, by a sale-and-lease-back arrangement, see Kim/Lewellen/McConnell (1978).

3) Here it is assumed, that the old bondholders' claims are not completely safe.

4) See Galai/Masulis (1976), pp. 62-64; Jensen/Meckling (1976), pp.334 -337; Smith/Warner (1979), pp. 118-119; Drukarczyk (1981), p. 309; Golbe (1981); Zechner (1982), pp. 187-189; Gavish/Kalay (1983), pp. 23-27; Green (1984), pp. 117-124. An explicit incorporation of taxes appears in Green/Talmor (1985).

stituting high-risk for low-risk-projects. Suppose, for example, that there are only two possible states s_1 and s_2 in t + 1 and that the originally implemented investment program yields the two possible outcomes $X(s_1) > F > X(s_2)$, wherein F denotes the bondholders' claims. Let there be an alternative investment program, which yields the two possible outcomes $Y(s_1) > X(s_1)$ and $Y(s_2) < X(s_2)$ while leaving the expectation constant. In this case, more risk is expressed in enlarging the outcome difference between the two states. If it is now possible to substitute the $Y(\cdot)$- for the $X(\cdot)$-program, stockholders would fully reap the gain $Y(s_1) - X(s_1)$ at the occurrence of s_1 while creditors would completely bear the loss $X(s_2) - Y(s_2)$ at the occurrence of s_2. Therefore, the substitution will be profitable for stockholders and this may occur even in those cases where the $Y(\cdot)$-program has a lower total market value than the $X(\cdot)$-program.

(b4) The Underinvestment Problem[1)]

The existence of risky debt claims may render investment projects unprofitable for stockholders although the projects may have a positive net present value (NPV). The reason for the occurrence of this case lies again in the sharing rule properties of the debt contract mentioned in (b3). Suppose stockholders raise new equity to finance a positive NPV-project. If the creditors' claims are not completely safe, i.e., if there are at least some future states where the debt claims exceed the surplus of the already given investment program, the outcomes of the new project in these states first of all accrue to bondholders. The present value of the remaining outcomes for stockholders may thus be lower than the investment outlays and under these circumstances it does not pay for stockholders to put up new equity capital for the realization of a new project.

In the above description each agency problem has been presented in isolation to clarify the source from which it may emerge. However, if it is recognized that stockholders in most cases have simultaneously all possibilities studied above, the situation becomes more complicated. The debt financing of a new investment project, for example, can be interpreted as a combination of problems (b2) and (b4): First the project is financed by the issuance of new equity and an underinvestment problem may arise. Then a debt financed dividend payment is made at the rate of that part of the investment outlays which should be finan-

---

1) See Myers (1977); Gupta (1982).

ced with new debt. Due to the effects mentioned in (b2), stockholders by this means are able to capture more of the project's NPV. This has led some researchers to the hypothesis that the underinvestment problem may be mitigated through the issuance of new debt.[1] That is, however, only one possible case. The author has shown elsewhere (see Ewert (1984); Ewert (1986a), pp. 195-226), that the debt financing of new projects may lead to overinvestment as well and, more surprisingly, even to more underinvestment than in the case of pure equity financing[2]. An equal spectrum of possibilities emerges if it is assumed that new projects may be financed at least partly by the proceeds of a liquidation of existing assets. This problem results in a combination of (b1), (b3) and (b4), but includes a somewhat different kind of asset substitution as (b3), because the liquidation of the existing program is not a pre-condition for the realization of the new one (in (b3) it was assumed that the programs considered are *alternatives*). For an elaboration of the over- and underinvestment problems in the case of liquidation-financing and for an integration of debt-and-liquidation-financing cases the reader is referred to Ewert (1986a), pp. 227-245.
A main feature of the above mentioned overinvestment situations is the (sometimes sharp) decline in the value of the bondholders' claims: An overinvestment can only be profitable for stockholders if they do not bear the NPV-diminution but this burden is transferred to creditors.

The agency problems so far described emerge from the structure of the financial claims against the firm. Occasionally, some problems of informational asymmetry are also included in a list of agency problems (see, for example, Barnea/Haugen/Senbet (1985), p. 38). Within this article that approach is not adopted because it would ultimately result in incorporating nearly all problems which are studied in several branches of the information economics literature. The goal of this paper (and of most papers in the financial agency literature the author knows) is more modest and lies in studying the problems resulting merely from the financial structure. However, integrating the aspects of informational asymmetry may be a fruitful and very exciting task

---

1) See Myers (1977), pp. 165-166; Smith/Warner (1979), p. 137; Stulz/Johnson (1985), p. 515.

2) See Ewert (1985), for an analysis of some determining factors of this agency problem.

for future research.[1)]

2.3 A Framework for an Investigation of Covenants

To identify the ultimate bearer of the welfare losses resulting from agency problems and to establish a setting in which covenants including accounting systems can be studied, the following premises are set:

P.1: Agency problems of equity are assumed to be not existing. At any time the managers act in the stockholders' interest.

P.2: The participants in the market are assumed to be able to make rational and unbiased estimates of the financial consequences from whatever policy the managers choose.

P.3: Contracts which directly and definitely specify a certain firm policy for all time-state-combinations are assumed to be not possible.

P.4: The stockholders and bondholders of the firm considered are assumed to be different, so that conflicts of interest may occur.

P.5: The structure of the capital market allows the representation of individual preferences by the with-dividend market value of individual investment holdings (for example, there is sufficient "competitivity" and spanning in the capital market).

P.6: By forces of a competitive capital market, individuals who become new claimants of a firm at any time only get the risk-equivalent return of their investment.

P.7: Possible market control mechanisms of agency problems (for example, take-overs, recontracting of financial claims, side-payments, etc.) are assumed to be not fully effective.

The above scenario characterizes most of the literature on debt-related agency problems and its assumptions may be discussed and justified in several ways (see for a detailed discussion Ewert (1986a), pp. 25-49). Within this framework, it can easily be shown, that the initial stockholders ultimately bear all wealth losses resulting from agency problems and thus have an incentive to engage in actions to mitigate the agency problems. The proof of this statement proceeds as

---

1) First steps are made by John/Nachman (1985), who investigate the underinvestment problem within a sequential rational expectations equilibrium, wherein stockholders may build a reputation.

follows [1] (all symbols in this section relate to the time of incorporation of the firm): Due to P.4 and P.5, the initial stockholders maximize the with-dividend value of the equity E(p) from any policy p (i.e. a flexible plan which determines the investment and financing program for all time-state-combinations):

$$E(p) = A(p) + S(p) \rightarrow \max! \qquad (1)$$

where A(p) denotes the initial dividend and S(p) the present value of the shares of policy p respectively. Using the cashflow identity (total cash inflows = total cash outflows) A(p) may be expressed as:

$$A(p) = D(p) - I(p) \qquad (2)$$

where I(p) denotes the investment outlays and D(p) the funds raised from the initial creditors (note, that there are no operating cash flows at the time of incorporation and that the initial dividend A(p) may be negative, which corresponds to the initial stockholders' part of financing I(p)). Furthermore, S(p) can be written as:

$$S(p) = V(p) - \tilde{D}(p) \qquad (3)$$

where V(p) denotes the present value of the firm as a whole and $\tilde{D}(p)$ equals the present value of the initial creditors' claims. Using (2) and (3), (1) can be written as:

$$E(p) = V(p) - I(p) - (\tilde{D}(p) - D(p)) \rightarrow \max! \qquad (4)$$

Due to P.2 and P.6, $\tilde{D}(p)$ equals D(p) and (4) reduces to:

$$E(p) = V(p) - I(p) = NPV(p) \rightarrow \max! \qquad (5)$$

where NPV(p) denotes the net present value of the firm as a whole when policy p is implemented. Let now be p* that policy which solves (5) in the absence of any agency problems. Those problems (as described in 2.2), however, will lead to a deviation from p* after the initial funds have been raised. The market participants anticipate the policy which will be optimal for future stockholders (P.1, P.2, P.4, P.5) and price their claims accordingly, so that NPV(p*) is no longer attainable, because possible market control mechanisms are neither fully effective (P.7). Furthermore, p* cannot be directly specified in a contract due to P.3. Therefore, the market anticipates a policy $\tilde{p}$, which will be influenced by agency problems and from the definition of p* and the

---

1) The following line of reasoning is performed similarly in Smith/Warner (1979), pp. 157-159; Smith (1980), pp. 342-345; Ewert (1986a), pp. 53-64.

above discussion it follows that $E(\tilde{p}) = NPV(\tilde{p}) \leqq NPV(p^*) = E(p^*)$ must be true,[1] so that the initial stockholders suffer all wealth losses resulting from agency problems.

If the use of debt principally increases the NPV (for example due to tax considerations) and the initial stockholders want to reap these benefits, they have somehow to convince the market that the forecasted agency problems (which ultimately lower the debt advantages) will not occur. Thus the initial stockholders are interested in implementing reliable mechanisms - henceforth called restrictions and symbolized by r - to mitigate the agency problems. Stockholders may, for example, restrict their future action space by means of dividend and financing constraints relying on accounting numbers, collateralization of assets, etc. ... Using restrictions, the realized policy may in general be expressed as a function $p(r)$. However, the use of restrictions regularly involves some cost (for example, cost in specifying and auditing accounting systems), which in general depends on the restriction and on the policy chosen. This cost is symbolized by $k(r;p(r))$. Therefore, the stockholders' problem is to determine that restriction r^*, which solves the following program:

$$NPV(p(r);k(r;p(r))) \rightarrow \max_{r \in R}! \qquad (6)$$

where R denotes all feasible combinations of restriction parameters, and the set R itself may be constrained by law (see Ewert (1986a), pp. 61-64 for a detailed discussion). Obviously $NPV(r^*)$ can never exceed $NPV(p^*)$ and with a suitable definition of R it follows that $NPV(r^*) \geqq$ $\geqq NPV(\tilde{p})$ must hold.[2]

This scenario constitutes a framework in which the role of bond covenants may be studied. By inspection of (5) and (6) such covenants are

---

1) The difference $NPV(p^*)-NPV(\tilde{p})$ depends on the possibilities to mitigate agency problems by the use of special financial contracts (call provisions, convertible debt, etc.). See for a discussion of these subjects Bodie/Taggart (1978); Barnea/Haugen/Senbet (1980); Haugen/Senbet (1981); Green (1984); Kudla (1984); Barnea/Haugen/Senbet (1985), pp. 85-111. In the following it is assumed that specially designed financial contracts do not fully resolve the agency problems.

2) The set R may be defined in such a way that it always contains the "zero-restriction", i.e., stockholders choose not to impose any restriction on their future action space and by this means the policy $\tilde{p}$ is always attainable.

advantageous if they do not allow great deviations from the value-maximizing-policy p* and at the same time are not too costly to be implemented. The above discussion has shown that *the initial shareholders* regularly are interested in some kind of creditor protection, as they gain from such protection through a higher price for the bonds and/or a lower rate of interest. Furthermore, the type of creditor protection in the agency framework has an intuitive appealing content: Prevent those actions, which would be taken and/or omitted by the stockholders after the debt issuance only because of wealth-transfer aspects.

Before proceeding to the next section, the author wants to emphasize an important remark: It is true that (6) principally describes a framework for an analysis of bond covenants but a detailed development of the optimal restriction-parameter-*portfolio* for a given firm in a given situation turns out to be very complicated. The author is not aware of any study which has already accomplished this very demanding task. Rather the existing studies (including those of the author) only carry out partial analyses in that they investigate the efficacy of some a priori selected types of covenants (for example, security agreements, accounting-based payout constraints, etc.)[1] For each type of covenant the consequences with regard to agency problems are studied so that positive and negative features of the selected type of covenant can be identified. With these results one can - strictly speaking - only form an *idea* of why the selected constraints might have been chosen in real situations and/or of what might be the survival features of existing contracts. Therefore, the third task of the FTA mentioned in 2.1 has not yet been fully accomplished. This fact has to be carefully taken into consideration in interpreting empirical results and/or in assessing the usefulness of predictions derived from the FTA in the field of positive and normative accounting research.

3. The General Relationship Between the FTA and External Accounting

3.1 Positive and Normative Questions in Accounting

The distinction between positive and normative questions in accounting follows the usual distinction between positive and normative theories in economic science. Positive accounting research concerns itself with an explanation of observed phenomena in the field of accounting. Typical problems in this branch of accounting research are, for example, how the use of accounting numbers in bond covenants might be explained,

---

1) Barnea/Haugen/Senbet (1985), p. 39, also emphasize this point.

why managers voluntarily choose some accounting methods and others not if they have discretion to do so etc. On the other hand normative accounting research deals with the question of how an accounting system *should* be. Those questions characterize particularly the viewpoint of the legislator, who wants to determine an accounting system to achieve some previously selected goals (for example "true and fair view", protection of bondholders' interests, protection of minority shareholders' interests, etc.). It has to be emphasized that the main part of the accounting literature in West Germany deals with normative questions.[1)]

There is an important link between positive and normative accounting research. Suppose the legislator wants to achieve a certain goal by the use of an accounting system. To determine an accounting system being suitable for the achievement of this goal the legislator has to develop hypotheses about the efficacy of alternative accounting measurement rules, that is, to develop and/or use theories about the consequences of alternative accounting systems with respect to the previously selected goal. To guarantee that the theories used in this procedure have something to do with reality the hypotheses should among other things include knowledge of positive accounting theories, which have proven to be useful in explaining observed phenomena. Thus a theory, which explicitly deals with problems relevant to the legislator and whose testable implications have not yet been refused during the process of empirical testing (i.e., the evidence is not substantially inconsistent with the implications and/or predictions derived by the theory) turns out to be a good tool for deriving the desired hypotheses. Therefore the results of positive accounting research may provide a great part of the knowledge which is necessary to answer normative questions in accounting.[2)]

3.2 The FTA and Normative Accounting Research

The FTA as shown in section 2 constitutes a framework which may be used to study the role and functioning of bond covenants, i.e., to identify mechanisms which are suitable to protect the claims of bondholders in the sense described in 2.3. As stated in the introduction, it is one of the major goals of the accounting system regulated by law in West

---

1) With respect to empirical accounting research in West Germany see the reviews by Coenenberg et al. (1978); Coenenberg/Möller/Schmidt (1984).

2) Jensen (1983), pp. 320-323, also emphasizes this argument.

Germany to protect creditors' interests (pci). Accordingly there are obvious relations between these two subjects.

First of all the FTA enables to interpret the goal "pci" as a vehicle to attain more allocative efficiency: Without any restrictions, creditors could be harmed by shareholders by deviations from the overall-value-maximizing policy. Preventing such deviations (see relation (6)) implies to implement a policy which will be near the overall-value-maximizing policy and thus leads to more allocative efficiency. Note, however, that this interpretation does not at all imply that the creditor protection has to be achieved by the intervention of the *legislator*. The FTA makes no statements regarding the necessity of regulatory interventions.[1] It rather makes hypotheses about the likely contracting outcomes which can be expected in private markets. Therefore the *optimality concept* given in (6) should not be overstated if the FTA is used to analyze accounting systems regulated by law, because a major problem in such an environment remains unresolved, namely: Why is regulation necessary? Furthermore, there are regularly many goals which at least the legislator in West Germany wants to achieve simultaneously by the use of an accounting system,[2] so that an optimality concept based on just one goal would not be suitable for the law-maker (this conclusion does not change if the agency problems of external equity would have been included in the framework of section 2.3).

That the *optimality concept* described in (6) is not directly applicable for the legislator does, however, not imply that the FTA is useless for regulatory problems. On the contrary: It is true that the goal "pci" is not the only objective to be achieved by regulatory interventions but it is recognized as a very important one in West Germany by all parties. Therefore the decisions of the legislator may be improved by the use of a framework which allows to provide precise and in some cases even unknown knowledge about the consequences of an accounting system with regard to the "pci"-objective.[3] This is exactly the way in which the FTA may fruitfully be used.

---

1) Such interventions are often justified by arguments of market failure with respect to the market for accounting information. A very illuminating discussion of potential pitfalls in these arguments is contained in Leftwich (1980). See for a more recent discussion of regulatory problems in financial accounting Schildbach (1986), pp. 89-98.

2) See, for example, Baetge (1976); Ellerich (1986).

3) A similar argument regarding the use of various research strategies for accounting policy decisions appears in May/Sundem (1976).

It has to be recognized, that the wealth of any group of claimants depends on the investment and financing policy chosen by the firm. Thus if a particular constraint shall be designed to improve the position of creditors one has first to develop hypotheses about the change in the investment and financing policy which will occur after the constraint has been implemented. From this analysis the change in the bondholders' position can be derived. This shows that a theory of firm policy serves as a necessary prerequisite for statements regarding the usefulness of particular accounting standards to achieve the "pci"-goal. As to the author's view, for example, it is not sufficient to justify the usefulness of the well-known "principle of caution" for the "pci"-objective simply with the widespread argument that more funds are retained in the firm. Rather it has to be analyzed under what circumstances and in which way the "principle of caution" influences the investment and financing policy and the author has shown elsewhere, that such an analysis may lead to conclusions which are very different from the already existing opinions.[1)]

As described in section 2 the FTA fully incorporates the incentives of stockholders to harm creditors as well as the interdepencies between firm policy and restrictions. It deals explicitly with problems which are also relevant to the legislator and thus meets in any case the first of the two usefulness-conditions mentioned at the end of 3.1. The second condition refers to the compatibility of the theory's implications with empirical data. The way in which the FTA fulfills this second condition is described in the following paragraph.

3.3 The FTA and Positive Accounting Research

As shown in 3.1 positive theories try to explain observed phenomena. In the ideal case the definitions and assumptions of the theory are used to obtain implications which are not immediately obvious by simply observing the initial premises. As far as the implications concern observable phenomena they can be confronted with the empirical evidence. If the data is considerably inconsistent with the implications the theory has to be modified, otherwise it temporarily may be accepted as an explanation of reality.

The above arguments describe an ideal process, which the FTA has not yet fully accomplished. The explanation process in the FTA proceeds in a slightly different way and may be characterized as follows:

---

1) See Ewert (1986a).

Usually some real contracts are observed (for example, bond covenants which include investment-, financing- and/or payout constraints based on accounting numbers, security agreements, etc.). These contracts then are integrated into agency models to identify positive and/or negative features with respect to resolving the agency problems. If there are enough positive features of a contract to be derived from the analysis the observation of the contract in reality is said to be consistent with the hypothesis that the contracting parties try to mitigate agency problems. As far as the theory enables to make more concrete hypotheses regarding specific situations (for example, the higher is the debt-equity ratio the tighter the payout constraint should be), these implications often are tested as well. This may lead to a further confirmation of the theory if the data is consistent with the hypotheses. These explanations then are partly used in some branches of the positive accounting literature to perform further tests. The "economic consequences"-approach,[1] for example, takes as given restrictions based on accounting numbers and their explanation derived from the FTA. Its goal is to explain the voluntary choice of accounting techniques by managers if the contracts give them discretion to do so as well as the voting or lobbying behavior of firms with respect to mandatory accounting rule changes. The hypotheses are derived from arguments of costly contracting which are similarly used in the FTA. If the empirical evidence is consistent with the hypotheses this leads not only to a temporary acceptance of the hypotheses and the "economic consequences" approach itself, but also to a further confirmation of the FTA from which this approach has partly been derived.

In view of the fact that up to now positive research in the agency literature takes the form described above one can say that a great part of the empirical evidence is consistent with the implications of the FTA.[2] This theory enables a meaningful interpretation of many elements in financial contracts the function of which was previously not well

---

1) See for a review Holthausen/Leftwich (1983).

2) See for studies regarding the explanation of observed contracts Smith/Warner (1979); Smith (1980); John/Kalay (1982), pp. 467-468; Kalay (1982); Leftwich (1983). With respect to the "economic-consequences" approach see Deakin (1979); Dhaliwal (1980); Bowen/Noreen/Lacey (1981); Leftwich (1981); Lilien/Pastena (1982); Daley/Vigeland (1983); Larcker/Revsine (1983); Lys (1984); Healy (1985); Zimmer (1986). See for empirical studies regarding the explanation of observed financing structures Thatcher (1985) and Kim/Sorensen (1986). Empirical work with respect to possible wealth transfers in processes of financial restructuring is reviewed by Jensen/Smith (1985), pp. 112-117.

understood. The FTA allows, for example, an explanation of the widespread use of the so called "financing rules" based on accounting numbers,[1] whereas such an explanation has sometimes been described as scientificly not possible.[2] With respect to positive accounting research and the "economic consequences"-approach, which is derived from the agency literature, Holthausen/Leftwich (1983), p. 79, even characterize this approach as "the most innovative and promising in financial accounting". Therefore, the second of the two usefulness conditions mentioned at the end of 3.1 seems to be fulfilled by the FTA as well.

Remember, however, the arguments presented at the end of 2.3, where it was argued, that the FTA up to now has not fully accomplished the task of deriving an optimal restriction-*portfolio* in a given situation. This fact may have important consequences with respect to the interpretation of positive research in the agency literature for the following reasons: In view of the ideal explanation-process described at the beginning of this paragraph, an explanation of observed contracts would mean that the contracts have to be fully endogenized by the theory. This amounts to an analysis, in which first of all it is shown, that from a given set of premises an optimal set of restriction parameters emerges. These predictions have to be confronted with reality in a second step, i.e., contracts observed in a real situation which corresponds to a set of assumptions in theory have to be compared with contracts that theory would predict for this situation. Since the extant FTA does not fully endogenize contracts, the results from positive research in agency literature have therefore to be interpreted with caution. There are, for example, several different mechanisms which have positive features with respect to resolving the underinvestment-problem, especially several types of payout constraints and - as was recently shown by Stulz/Johnson (1985) - secured debt.[3] If it is recognized, that both mechanisms are costly to be implemented, the theory not only has to identify positive and negative features, but it also has to show under which circumstances and in which extent it is optimal to use only one or both or even none of the above mentioned constraints. Furthermore

---

1) See, for example, Kalay (1982), pp. 218-219, pp. 228-232; Leftwich (1983); Ewert (1984).

2) See Bieg (1983), p. 496.

3) See for other positive features of secured debt with respect to the mitigation of agency problems Smith/Warner (1979), pp. 127-128; Swoboda (1982); Rudolph (1982); Drukarczyk (1983); Schildbach (1983); Rudolph (1984); Rudolph (1986).

such an analysis has to consider other restriction parameters and other possible agency problems as well. Therefore the explanations so far obtained from positive research in agency literature are very preliminary in character and a further development of theory is necessary for an improvement on this point.[1)] Within the current state of knowledge, however, the FTA seems to be very fruitful in explaining observed phenomena.

4. Some Applications of the FTA in the Field of External Accounting

4.1 Preliminaries

While the previous sections were concentrated on general topics the remainder of the paper presents some results of existing models. Due to space limitations the proofs of the stated results are generally omitted and only the key ideas are outlined. For a more detailed discussion the reader is referred to the literature cited below.

The analysis below is performed under the general assumptions mentioned in 2.3 and within the following simplified framework: A two-period-three-date (t_0, t_1, t_2) model will be considered. The initial investment program at time t_0 has already been fixed and the involved initial investment outlays have been financed partially by the issuance of a pure discount bond with nominal claims F maturing at t_2. The t_0-program yields in t_1 and t_2 positive but uncertain outcomes. To concentrate only on wealth-transfer aspects taxes are not explicitly incorporated into the analysis and bankruptcy costs at the date of maturity t_2 are assumed not to exist.

The model now explicitly studies the stockholders' optimal policy in t_1, where the firm can invest any amount in a new project, whose outcome function in t_2 is assumed to be state-dependent, strictly concave with respect to the investment outlays and to have strictly positive marginal outcomes for all investment levels, so that the total cash flows from the new project in t_2 are positive for positive investment levels. Furthermore, the present-value- and the NPV-function of the new project are - within a usual state-preference valuation approach - strictly concave as well. Except for the possibility of investing in t_1 stockholders may either be allowed to issue new debt of at least the same priority as the old one and to liquidate the already existing program in t_1 (henceforth called "complex scenario") or be not allowed to do so (henceforth called "simple scenario"). Thus in a simple scenario

---

1) A similar argument appears in Raviv (1985).

the investment outlays in t_1 have to be financed exclusively by the retention of operating cash flows and/or the issuance of new equity and/or the issuance of new (but subordinated) debt (all three methods of financing are, however, equivalent under the above assumptions). Furthermore, the agency problems (b1), (b2) and (b3) stated in 2.2 are not possible in a simple scenario.[1] A similar argument holds for the combined agency problems described in 2.2, so that the set of possible agency problems in a simple scenario reduces to the underinvestment problem (see (b4) in 2.2). If all agency problems described in 2.2 are supposed to exist one has therefore to consider complex scenarios.

The results presented below are only a selection and do not cover the whole set of existing results with regard to the role of accounting systems within an agency framework. In what follows the term "restriction" or "constraint" always expresses a reduction in the action space for stockholders and an accounting system is viewed as a means of providing a direct payout constraint[2], i. e., the determination of an amount which serves as an upper limit (payout potential) for the distribution of funds to shareholders to mitigate the bondholder-stockholder conflict. This corresponds to the "payout allocation function" of the accounting system regulated by law in West Germany with respect to the "pci"-goal.

4.2 The Relationship Between a Mitigation of Agency Problems and the Goal "Protection of Creditors' Interests"

In paragraph 3.2 it was shown that although the *optimality concept* (6) of the FTA may not be suitable for a legislator, an analysis of agency problems should nevertheless be performed. To stress the importance of this argument one can show that the following theorem must be true:

Theorem 1: *If the implementation or tightening of a restriction r leads to a mitigation of agency problems in t_1 (that is, NPV $(r;t_1)$ exceeds NPV (t_1) without the implementation or tightening of r), the wealth of the initial bondholders in t_1 increases at least by the raise in NPV of the firm as a whole. Decreases in wealth of the initial bondholders in t_1 induced by a restriction are possible if and only if the implementation or tightening of a restriction r leads to a raise in agency problems (for a proof see Ewert (1986a), pp. 249 - 251).*

1) The asset substitution problem (b3) ultimately implies a liquidation of the already existing program and is therefore not possible in a simple scenario.

2) See for a more detailed discussion of several types of payout constraints Kalay (1982); Ewert (1986a), pp. 100-143.

The proof of Theorem 1 relies on the idea, that the policy p(r) chosen by stockholders *with* r (or with a tightening of r) was always possible for them *without* r. If there is $p(r) \neq \tilde{p}$ ($\tilde{p}$ denotes the stockholders' optimal policy without r) and if NPV $(p(r);t_1) >$ NPV $(\tilde{p};t_1)$ holds, the reason that stockholders did not choose p(r) before must be that more than the increase in NPV would be captured by the initial bondholders. This leads ultimately to Theorem 1, which gives strong arguments in favor of an agency analysis, if the purpose of the balance-sheet is - among other things - the "pci".

Furthermore, note that the contents of Theorem 1 are ultimately responsible for the fact, that at the initial date t_0 the initial stockholders fully bear all wealth losses resulting from agency problems and capture all gains from mitigating agency problems (as shown in 2.3): If a restriction r resolves agency problems, the initial bondholders anticipate the consequences stated in Theorem 1 and pay for the increased value of their claims, so that the initial stockholders are benefited.

4.3 The Role of the "Principle of Caution" in Resolving Agency Problems in Simple Scenarios

As far as the "payout allocation function" with respect to the "pci"-goal is concerned the well-known "principle of caution" seems to have received some aura of divinity in the accounting literature.[1] The essence of its justification lies in the simple fact mentioned in 3.2, i. e., the lower valuation of the firm's assets in the balance sheet leads to an at least temporarily retention of more funds in the firm and this is hypothesized to strengthen the firm's ability to repay debt. Even most recent publications dealing with the problem of how to achieve more "pci" by, among other things, the use of suitable accounting systems go - with respect to accounting measurement rules - not very far beyond stating the above hypothesis and promote an extension of the "principle of caution" over the already existing level.[2]

The FTA has proven to be very useful for an analysis of the efficacy of alternative accounting measurement rules because it has allowed not only to identify sufficient conditions under which the above arguments

---

1) In a commentary of the new balance-sheet-directive law in West Germany, Weber (1986), p. 125 describes the "principle of caution" as the most important accounting principle.

2) See, for example, Hemmerde (1985), pp. 397-443; Bitz/Hemmerde/Rausch (1986), pp. 186-203.

are true but also to discover scenarios in which they may be (but need not be) *false*. Unfortunately the conditions which lead to the former case seem to be no more realistic than those which lead to the latter one, so that the above arguments in favor of the "principle of caution" should be somewhat qualified. To justify this statement the sufficient conditions for the "positive" case have first to be examined.

It turns out that the simple scenario of the model described in 4.1 is sufficient for all hypothesized positive features of the "principle of caution" with respect to the "pci"-goal. Remember that in a simple scenario only the underinvestment problem is possible to exist at t_1. A natural means of mitigating this problem is to establish a minimum investment constraint at t_1, i. e., to determine an investment level which stockholders are not allowed to fall short of. A direct payout constraint based on an accounting system provides such a minimum investment constraint for the following reasons: Let $NE(r;t_1)$ denote the net earnings of accounting system r (all symbols refer to t_1) and let $X(t_1)$ be the operating cash flow of the t_0-program. $NE(r;t_1)$ now may be expressed as:

$$NE(r;t_1) = X(t_1) - \Delta(r;t_1) \qquad (7)$$

where $\Delta(r;t_1)$ denotes all differences between $X(t_1)$ and both the expenditures and the earnings of accounting system r (for example, depreciation, etc.). It is assumed that $\Delta(r;t_1) > 0$. If $B(t_1)$ denotes cash inflows from new equity and $DS(t_1)$ denotes cash inflows from the issuance of new subordinated debt, the following cash flow identity must hold for t_1 in the simple scenario of the model described in 4.1:

$$A(t_1) + I(t_1) = X(t_1) + B(t_1) + DS(t_1) \qquad (8)$$

where $A(t_1)$ and $I(t_1)$ denote the payouts to stockholders and investment outlays respectively (note that there are no payments of interest and no principal payments in t_1). Since $B(t_1)$ and $DS(t_1)$ have no wealth-transfer potential their use for the payment of funds to stockholders need not be restricted. Therefore the following payout constraint is assumed:

$$A(t_1) \leqq \max \; \{NE(r;t_1);0\} + B(t_1) + DS(t_1) \qquad (9)$$

where the max $\{\cdot\}$-operator means that stockholders are not obliged to compensate for negative net earnings by issuing new equity and/or subordinated debt (note that $A(t_1) \geqq 0$ as a negative payment to stockholders implies ultimately the payment of new equity into the firm, for which there is an own variable $B(t_1)$). By using (7) and (8), (9) may be expressed as:

$$I(t_1) \geqq \min \{\Delta(r;t_1);X(t_1)\} \qquad (10)$$

Inequality (10) gives the desired result and shows the minimum investment constraint implied by the direct payout constraint (9). The "principle of caution" leads to lower asset values and thus increases the minimum investment constraint by an increase in $\Delta(r;t_1)$. The detailed analysis of the efficacy of increased minimum investment constraints in a simple scenario (see John/Kalay (1982); Ewert (1986a), pp. 148 - 180) can be summarized by the following theorem:

Theorem 2: Let there be a simple scenario of the model presented in 4.1. Then a restriction of the type described in (9) becomes the more relevant the more initial debt (measured by the claims F) is used. An increase in $\Delta(r;t_1)$ can never lead to a decrease in the stockholders' realized investment level and there may be states in t_1, where even a small increase in $\Delta(r;t_1)$ may lead to a relatively large increase in the realized investment. Furthermore, an increase in $\Delta(r;t_1)$ will never harm the bondholders, but it may induce overinvestment. Apart from cost considerations an increase in $\Delta(r;t_1)$ may become problematical from an overall-value-maximizing-viewpoint only if it leads to overinvestment in at least one t_1-state.

Theorem 2 shows, that within a simple scenario the "principle of caution" has indeed absolutely positive features with respect to the "pci"-goal. It also shows, however, that this may be accompanied by overinvestment and thus by a non optimal investment policy. The simple scenario described in 4.1 may be interpreted as a basic debt contract in which the liquidation of assets is prohibited and either stockholders are not allowed to issue any new debt or there are "me-first-rules" (see Fama/Miller (1972), pp. 151 - 152) in favor of the initial creditors. Especially the last two characteristics are rarely found in reality. Therefore it seems interesting to relax the financing assumptions of the simple scenario and to perform further analyses.

4.4 Payout Restrictions in Complex Scenarios

In this paragraph it is assumed that stockholders are able to liquidate the already existing program in t_1 and to issue new debt with the same priority as the old one. In this case it can be shown that it is optimal for stockholders always to use at least one of these two activities (regularly a combination of both) in the maximum amount they are allowed to (see Ewert (1986a), pp. 239 - 249), whereas financing instruments without any wealth transfer potential are inferior for them. Furthermore the combined agency problems described in 2.2 may arise.

In such an environment a direct payout constraint of the type given by (9) may be viewed as providing a specific combination of both a minimum investment constraint and a financing restriction for the following reasons[1]: Let $L(t_1)$ denote the proceeds from the liquidation of already existing assets and let $DE(t_1)$ express the proceeds from the issuance of new debt with equal priority as the old one in t_1 respectively. In a complex scenario of the model described in 4.1 the cash-flow identity (8) now becomes:

$$A(t_1) + I(t_1) = X(t_1) + B(t_1) + L(t_1) + DS(t_1) + DE(t_1) \quad (11)$$

It is again assumed that payout restriction (9) holds and that $NE(r;t_1)$ may be expressed as in (7). Note that in principle in a complex scenario $\Delta(r;t_1)$ may depend on $L(t_1)$. In the following it is, however, assumed that the determination of $NE(r;t_1)$ occurs immediately *before* stockholders choose their optimal policy, so that $\Delta(r;t_1)$ does not depend on the stockholders' actions. Using (11) and (7), (9) may be expressed as:

$$L(t_1) + DE(t_1) \leqq I(t_1) - \min\{\Delta(r;t_1);X(t_1)\} \quad (12)$$

Inequality (12) may be interpreted as follows: Because of $L(t_1) \geqq 0$ and $DE(t_1) \geqq 0$, the right-hand side of (12) expresses a minimum investment constraint similar to the one given by (10). Moreover the sum of $L(t_1) + DE(t_1)$ may not go beyond that part of the t_1-investment which exceeds the amount given by the minimum investment constraint. This latter amount has thus to be financed by instruments without any wealth transfer potential. Therefore, a direct payout constraint can be interpreted as a combination of two in principle different restrictions. First it includes the following pure financing constraint:

$$L(t_1) + DE(t_1) = 0, \text{ if } I(t_1) \leqq \min\{\Delta(r;t_1);X(t_1)\} \quad (13)\text{a}$$

$$L(t_1) + DE(t_1) \leqq I(t_1) - \min\{\Delta(r;t_1);X(t_1)\}, \quad (13)\text{b}$$

$$\text{if } I(t_1) \geqq \min\{\Delta(r;t_1);X(t_1)\}$$

Note that (13)a and (13)b allow for $I(t_1) = 0$ and therefore do not contain a minimum investment constraint. If, however, (13)a and (13)b are combined with a pure minimum investment constraint as given by (10) one obtains exactly the relations expressed in (12).

It can be shown, that for an analysis of the efficacy of a direct payout constraint within a complex scenario the above splitting turns out to be very useful and the relative importance of the pure financing

---

1) See for a detailed discussion of this splitting Ewert (1986a), pp. 254-272.

constraint and the minimum investment constraint for the functioning of the payout restriction can be identified (for a detailed discussion see Ewert (1986a), pp. 254-346). Furthermore, the following theorem emerges from this analysis:

Theorem 3: *Let there be a complex scenario of the model described in 4.1. Then there may exist intervals of $\Delta(r;t_1)$, where an increase in $\Delta(r;t_1)$ induces such changes in the stockholders' optimal policy that the initial bondholders' wealth at t_1 decreases. Furthermore the direct payout constraint may be responsible for overinvestment even if $\Delta(r;t_1)$ does not exceed the overall value maximizing investment level.*

Theorem 3 shows that a more "cautious" valuation of the firm's assets (i.e., an increase in $\Delta(r;t_1)$) may actually harm the initial bondholders (for detailed examples see Ewert (1986a), pp. 293-316). If this case occurs it follows from theorem 1 that the increase in $\Delta(r;t_1)$ has increased agency problems as well. Thus the above analysis shows that the efficacy of the "principle of caution" depends on the scenario chosen. The complex scenario may be interpreted as a basic debt contract, which allows stockholders more discretion than does the basic debt contract of the simple scenario. Viewed in this way the above arguments emphasize the role of the restriction *portfolio* in which the payout constraint is embedded. As stated in paragraph 2.3 the detailed development of optimal restriction portfolios remains to be done by future research.

5. Concluding Remarks

As shown above the FTA proves to be a useful tool for an analysis of problems in financial accounting. By means of the FTA it can be understood that an accounting system may induce completely different consequences for different financing scenarios. These results imply that one should carefully specify the assumptions if an analysis of the functioning of an accounting system has to be performed (although such an explication of premises may be viewed as a natural prerequisite for every scientific analysis it unfortunately not always seems to be a naturalness in the field of accounting research).

The above analysis may and should be extended in several ways. Existing extensions of the model described in paragraph 4.1, for example, deal with the incorporation of investments in the financial markets[1] and

---

1) See Kim (1982) and Ewert (1986a), pp. 180-191, pp. 360-365.

the integration of agency models in a signaling framework.[1] Furthermore it is possible to identify many aspects of a payout constraint whose efficacy is not yet fully understood.[2] An example is the aspect of cumulativity (that is the possibility to build reservoirs of payable funds and to use them in later periods), whose effects with respect to agency problems are up to now only hypothesized[3] or studied within a simple scenario.[4] In any case agency theory opens up new perspectives for studying problems in financial accounting. This as well as the current research on information economics will hopefully enhance our knowledge about accounting phenomena.

References

Baetge, J. (1976): Rechnungslegungszwecke des aktienrechtlichen Jahresabschlusses, in: Bilanzfragen, Festschrift zum 65. Geburtstag von U. Leffson, Eds.: J. Baetge; A. Moxter; D. Schneider, Düsseldorf, 11-30.

Barnea, A.; Haugen, R.A.; Senbet, L. W. (1980): A Rationale for Debt Maturity Structure and Call Provisions in the Agency Theoretic Framework, in: JoF, vol. 35, 1223-1234.

Barnea, A.; Haugen, R.A.; Senbet, L. W. (1981): Market Imperfections, Agency Problems, and Capital Structure, in: Financial Management, vol. 10, 7-22.

Barnea, A.; Haugen, R.A.; Senbet, L.W. (1985): Agency-Problems and Financial Contracting, Englewood-Cliffs.

Bieg, H. (1983): Finanzierungsregeln, in: WiSt, vol. 12, 491-496.

Bitz, M.; Hemmerde, W.; Rausch,W. (1986): Gesetzliche Regelungen und Reformvorschläge zum Gläubigerschutz - Eine ökonomische Analyse, Berlin/Heidelberg/New York/Tokyo.

Black, F. (1976): The Dividend Puzzle, in: Journal of portfolio management, vol. 2, 5-8.

Black, F.; Scholes, M. (1973): The Pricing of Options and Corporate Liabilities, in: Journal of Political Economy, vol. 81, 637-654.

Bodie, Z.; Taggart, R.A. (1978): Future Investment Opportunities And The Value Of The Call Provision On A Bond, in: JoF, vol. 33, 1187-1200.

Bowen, R.M.; Noreen, E.W.; Lacey, J.M.(1981): Determinants Of The Corporate Decision To Capitalize Interest, in: JoAE, vol. 3, 151-179.

Coenenberg, A.G. et al. (1978): Empirische Bilanzforschung in Deutschland, in: Die Betriebswirtschaft, vol. 38, 495-507.

---

1) See John/Kalay (1985).
2) See Kalay (1982); Ewert (1986a), pp. 100-143.
3) See Kalay (1982), pp. 226-227.
4) See Ewert (1986b).

Coenenberg, A.G.; Möller, H.P.; Schmidt, F. (1984): Empirical Research in Financial Accounting in Germany, Austria and Switzerland - A Review, in: European Contributions to Accounting Research, Eds.: A.G. Hopwood, H. Schreuder, Amsterdam, 61-81.

Daley, L.A.; Vigeland, R.L. (1983): The Effects Of Debt Covenants And Political Costs On The Choice Of Accounting Methods, in: JoAE, vol. 5, 195-211.

Deakin, E.B. (1979): An Analysis of Differences Between Non-Major Oil Firms Using Successful Efforts and Full Cost Methods, in: The Accounting Review, vol. 54, 722-734.

Dhaliwal, D.S. (1980): The Effect of the Firm's Capital Structure on the Choice of Accounting Methods, in: The Accounting Review, vol. 55, 78-84.

Drukarczyk, J. (1981): Verschuldung, Konkursrisiko, Kreditverträge und Marktwert von Aktiengesellschaften, in: Kredit und Kapital, vol. 14, 287-319.

Drukarczyk, J. (1983): Kreditverträge, Mobiliarsicherheiten und Vorschläge zu ihrer Reform im Insolvenzrecht, in: ZfB, vol. 53, 328-349.

Ellerich, M. (1986): Zwecke des handelsrechtlichen Jahresabschlusses, in: Küting/Weber (Eds.): Handbuch der Rechnungslegung, Stuttgart, 155-160.

Ewert, R. (1984): Zur Beziehung zwischen Investitionsvolumen, Fremdfinanzierung und Bilanzkennzahlen, in: ZfbF, vol. 36, 825-841.

Ewert, R. (1985): Die Finanzierung zusätzlicher Investitionsprojekte mit Fremdkapital: Analyse eines kombinierten Agency Problems und dessen Bestimmungsfaktoren, in: Geld, Banken und Versicherungen, Eds.: H. Göppl, R. Henn, Karlsruhe, 1131-1144.

Ewert, R. (1986a): Rechnungslegung, Gläubigerschutz und Agency-Probleme, Wiesbaden.

Ewert, R. (1986b): Bilanztheorie und finanzierungsbedingte Agency-Costs: Grundsätzliche Beziehungen und ein Modell zur Analyse kumulativer Ausschüttungsrestriktionen, unpublished working paper, University of Passau.

Fama, E.F.; Miller, M.H. (1972): The Theory of Finance, New York.

Fama, E.F. (1978): The Effects Of A Firm's Investment and Financing Decisions on The Welfare Of Its Security Holders, in: AER, vol. 68, 272-284.

Galai, D.; Masulis, R.W. (1976): The option pricing model and the risk factor of stock, in: JoFE, vol. 3, 53-81.

Gavish, B.; Kalay, A. (1983): On the Asset Substitution Problem, in: JoFQA, vol. 18, 21-30.

Gjesdal, F. (1981): Accounting for Stewardship, in: Journal of Accounting Research, vol. 19, 208-231.

Golbe, D.L. (1981): The effects of imminent bankrupty on stockholder risk preferences and behaviour, in: Bell Journal of Economics, vol. 12, 321-328.

Green, R.C. (1984): Investment Incentives, Debt, and Warrants, in: JoFE, vol. 13, 115-136.

Green, R.C.; Talmor, E. (1985): The Structure and Incentive Effects of Corporate Tax Liabilities, in: JoF, vol. 40, 1095-1114.

Gupta, K. (1982): Determinants of Corporate Borrowing: A Note, in: JoFE, vol. 10, 115-116.

Hakansson, N.H. (1978): Where we are in Accounting: A Review of "Statement on Accounting Theory and Theory Acceptance", in: The Accounting Review, vol. 53, 717-725.

Haugen, R.A.; Senbet, L.W. (1981): Resolving the Agency Problems of External Capital through Options, in: JoF, vol. 36, 629-647.

Healy, P.M. (1985): The Effect of Bonus Schemes on Accounting Decisions, in: JoAE, vol. 7, 85-107.

Hemmerde, W. (1985): Insolvenzrisiko und Gläubigerschutz - Eine ökonomische Analyse gesetzlicher Normen zur Kapitalaufbringung und Kapitalerhaltung, Thun/Frankfurt am Main.

Holthausen, R.W.; Leftwich, R.W. (1983): The Economic Consequences of Accounting Choice, - Implications of Costly Contracting and Monitoring, in: JoAE, vol. 5, 77-117.

Jensen, M.C. (1983): Organization Theory and Methodology, in: The Accounting Review, vol. 58, 319-339.

Jensen, M.C.; Meckling, W.H. (1976): Theory of The Firm: Managerial Behaviour, Agency Costs And Ownership Structure, in: JoFE, vol. 3, 305-360.

Jensen, M.C.; Smith, C.W. (1985): Stockholder, Manager and Creditor Interests:Applications of Agency Theory, in: Recent Advances in Corporate Finance, Eds.: E.I. Altman, M.G. Subrahmanyam, Homewood, 93-131.

John, K.; Kalay, A. (1982): Costly Contracting and Optimal Payout Constraints, in: JoF, vol. 37, 457-470.

John, K.; Kalay, A. (1985): Informational Content of Optimal Debt Contracts, in: Recent Advances in Corporate Finance, Eds.: E.I. Altman, M.G. Subrahmanyam, Homewood, 133-161.

John, K.; Nachman,D.C. (1985): Risky Debt, Investment Incentives, and Reputation in a Sequential Equilibrium, in: JoF, vol. 40, 863-880.

Kalay, A. (1982): Stockholder-Bondholder Conflict And Dividend Constraints, in: JoFE, vol. 10, 211-233.

Kim, E.H. (1982): Discussion, in: JoF, vol. 37, 476-479.

Kim, E.H.; Lewellen, W.G.; McConnell, J.J. (1978): Sale-And-Leaseback Agreements And Enterprise Valuation, in: JoFQA, vol. 13, 871-883.

Kim, E.H.; McConnell, J.J.; Greenwood, P.R. (1977): Capital Structure Rearrangements And Me-First-Rules In An Efficient Capital Market, in: JoF, vol. 32, 789-810.

Kim, W.S.; Sorensen, E.H. (1986): Evidence on the Impact of the Agency Costs of Debt on Corporate Debt Policy, in: JoFQA, vol. 21, 131-144.

Krainer, R.F. (1977): Interest rates, leverage, and investor rationality in: JoFQA, vol. 12, 1-16.

Kudla, R.J. (1984): A Note on Resolving Agency Problems through Warrants, in: Journal of Business Finance & Accounting, vol. 11, 67-71.

Larcker, D.F.; Revsine, L. (1983): The Oil and Gas Accounting Controversy: An Analysis of Economic Consequences, in: The Accounting Review, vol. 58, 706-732.

Leftwich, R. (1980): Market Failure Fallacies And Accounting Information, in: JoAE, vol. 2, 193-211.

Leftwich, R. (1981): Evidence of the Impact of Mandatory Changes in Accounting Principles on Corporate Loan Agreements, in: JoAE, vol. 3, 3-36.

Leftwich, R. (1983): Accounting Information in Private Markets: Evidence from Private Lending Agreements, in: The Accounting Review, vol. 58, 23-42.

Lilien, S.;Pastena, V. (1982): Determinants of Intramethod Choice in the Oil and Gas Industry, in: JoAE, vol. 4, 145-170.

Lys, T. (1984): Mandated Accounting Changes and Debt Covenants: The Case of Oil and Gas Accounting, in: JoAE, vol. 6, 39-65.

May, R.G.: Sundem, G.L. (1976): Research for Accounting Policy: An Overview, in: The Accounting Review, vol. 51, 747-763.

Modigliani, F.; Miller M.H. (1958): The Cost of Capital, Corporation Finance and the Theory of Investment, in: AER, vol. 48, 261-297.

Myers, S.C. (1977): Determinants of Corporate Borrowing, in: JoFE, vol. 5, 147-175.

Raviv, A. (1985): Management Compensation And The Managerial Labor Market - An Overview, in: JoAE, vol. 7, 239-245.

Ross, S.A. (1973): The Economic Theory Of Agency: The Principal's Problem, in: AER, vol. 63, 134-139.

Rudolph, B. (1982): Können die Banken ihre Kreditsicherheiten vergessen?, in: Kredit und Kapital, vol. 15, 317-341.

Rudolph, B. (1984): Kreditsicherheiten als Instrumente zur Umverteilung und Begrenzung von Kreditrisiken, in: ZfbF, vol. 36, 16-43.

Rudolph, B. (1986): The Value of Security Agreements, in: Capital Market Equilibria,Eds.: G. Bamberg, K. Spremann, Berlin/Heidelberg/New York/Tokyo, 135-162.

Schall, L.D. (1972): Asset Valuation, Firm Investment And Firm Diversification, in: JoB, vol. 45, 11-28.

Schildbach, Th. (1983): Sicherheiten versus par conditio creditorum, in: Betriebs-Berater, vol. 38, 2129-2136.

Schildbach, Th. (1986): Jahresabschluß und Markt, Berlin/Heidelberg/New York/Tokyo.

Smith, C.W. (1980): On the Theory of Financial Contracting, in: Journal of Monetary Economics, vol. 6, 333-357.

Smith, C.W.; Warner, J.B. (1979): On Financial Contracting: An Analysis of Bond Covenants, in: JoFE, vol. 7, 117-161.

Stiglitz, J.E. (1969): A Re-Examination of the Modigliani-Miller-Theorem, in: AER, vol. 59, 784-793.

Stiglitz, J.E. (1974): On The Irrelevance of Corporate Financial Policy, in: AER, vol. 64, 851-866.

Stulz, R.M.; Johnson, H. (1985): An Analysis of Secured Debt, in: JoFE, vol. 14, 501-521.

Swoboda, P. (1982): The Relevance of Securing Debts, unpublished working paper, University of Graz.

Thatcher, J.S. (1985): The Choice of Call Provision Terms: Evidence of the Existence of Agency Costs of Debt, in: JoF, vol. 40, 549-561.

Titman, S. (1984): The Effect of Capital Structure on a Firm's Liquidation Decision, in: JoFE, vol. 13, 137-151.

Weber, H.K. (1986): Aussagefähigkeit des neuen Jahresabschlusses, in: Küting/Weber (Eds.): Handbuch der Rechnungslegung, Stuttgart, 123-153.

Zechner, J. (1982): Managerverhalten und die optimale Kapitalstruktur von Unternehmungen: Ein Überblick zur Theorie der Agency Costs, in: Journal für Betriebswirtschaft, vol. 32, 180-197.

Zimmer, I. (1986): Accounting For Interest By Real Estate Developers, in: JoAE, vol. 8, 37-51.

Asymmetric Information between Investors and Managers under the New German Accounting Legislation

Friedrich Fricke

Summary: This article analyses the ways in which the positions of investors and shareholders on the one side and of managers on the other side will be affected by the New Accounting Directives Law. Although the literature is emphasizing the many improvements resulting from the New Accounting Law, there are some consequences of miscellaneous character. In particular, the question arises in which way managers alter their attitude as a reaction to the law.

Table of Contents

1. Subject Limitations

For a variety of reasons, it is not possible to give full and comprehensive details on the subject of the Information Structure of the New German Accounting Legislation. The changes due to the Accounting Directives Law[1] are too wide ranging, commencing with classification requirements through valuation principles and up to the extended statutory audit and disclosure requirements of the financial statements. Starting in 1990, it will be compulsory to prepare consolidated balance sheets even on a worldwide basis which is a separate and voluminous

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

subject in itself.

The following subjects are therefore selected and discussed in this article:

- Possibilities of analysing the balance sheet,
- Possibilities of analysing the profit and loss account,
- Extended statutory audit requirements,
- Extended disclosure requirements.

Depending on their individual interests, the above-mentioned subjects affect the investor (principal) on the one hand and likewise the manager (agent) on the other hand.[2] This article is an attempt to show the effects of these changes resulting form the new law on the principal's and the agent's positions. In addition, a preliminary review is presented on how principal and agent will react to the reforms.

2. The Concept of "Information Structure"

It is difficult to define the concept of "Information Structure" clearly and completely. Due to the fact that this concept covers the entire Accounting Directives Law, the definition will be the following transcript of five properties:[3]

a) The concept "Information Structure" includes information media, information sources, parties interested in information and the contents of information.

b) The information media in the Accounting Directives Law are the balance sheet, the profit and loss account, the notes to the financial statements and the management report.

c) The information sources are the Federal Gazette and/or the financial statements filed with the Commercial Register (extended disclosure) and also the sources to which the investors have direct access (e.g. accounting records, agreements and credit lines amongst others).

d) Parties interested in information are today's and future principals (creditors and shareholders), agents (managers), the revenue authorities, and employees especially if they participate in the company's earnings.

e) The contents of information which in some cases have been improved by the new law concern, amongst others:

   - clearer and more understandable presentation due to classification formats, valuation provisions, statutory notes to the financial statement and management report for extended range of businesses obliged to set up financial statements, and particularly for companies,

- more dependable presentation because the valuation margin has been tightened and extended number of companies now being subject to annual audit of financial statements,
- accelerated preparation of the financial statements due to the requirement of a deadline to prepare and submit to the shareholders.[4)]

3. The Possibilities of Analysing the Balance Sheet

3.0 General Remarks

As far as the external balance sheet analysis is concerned, an enormous increase in the amount of items for analysis is to be expected. Due to the extension of disclosure, there will be about 300,000 limited liability companies whereas before there were 2,000 companies only which had to disclose according to stock corporation law or publicity law.

In future, the published financial statements of companies (limited liability companies, stock corporations and partnerships limited by shares) will provide a good insight into the various trade groups. It will, particularly for financial institutions,[5)] allow easier comparisons to be made between the companies of a branch. Thus, a wider range of people will be engaged in balance sheet analysing such as suppliers, customers or competitors who, up until now, had no access to these financial statements. The question arises whether or not new analysing methods have to be developed since the methods used to date were mainly concerned with the law classification formats[6)] of stock corporations.

Due to the new classification formats for the balance sheet (§§ 247, 266) as well as profit and loss account (§ 275), and due to the formal classification requirements (§§ 243 PP., 265, 275), the annual financial statements become clearer and more understandable. This is particularly true as far as periodical comparison is concerned.[7)]

3.1 Property Structure (Assets)

As far as the property structure is concerned, the information available to creditors will improve. Companies are now required to show their fixed assets gross (§ 268 Abs. 2). That is, historical acquisition/manufacturing costs must be shown as well as as the related accumulated depreciation.

The gross method gives a better view of the age structure of the assets; it shows the need to reinvest[8] and the extent of waste.

There are significant changes with respect to intangible fixed assets (patents, licences, goodwill, etc.) insofar as the intangible assets now have to be capitalized if they were acquired against payment (§ 248 Abs. 2). Combined with the information in the management report concerning research and development (§ 289 Abs. 2 Nr. 3), and with the help of time studies and intercompany comparisons, interesting conclusions about a firm's future prospects can be drawn. The information concerning the buying of patents and licences may even disclose a technological dependency.[9]

Medium sized and large companies have to disclose their company connections[10] by showing their shares in affiliated enterprises and participations (§ 266 Abs. 2). Notes to the financial statement must include specific disclosures concerning participations (20 % and more); in accordance with § 285 Nr. 11, name and seat of the affiliated enterprises, the percentage of shares of the capital, the equity and the results of the last financial year must be stated (a so-called list of participations). With the help of this information, there is also the possibility to look at the financial statements of the stated affiliated enterprises. In addition, loans, receivables and liabilities related to affiliated enterprises and/or participations must appear as special items (§ 266 Abs. 2). Contingent liabilities must be shown separately (§ 268 Abs. 7). All this provides a better view of the affiliations of enterprises. This is not irrelevant since a parent company in difficulties will nearly always affect a subsidiary.

An improvement concerning the liquid assets analysis[11] is that now companies will have to mention separately the amounts of receivables with a remaining term of more than one year (§ 268 Abs. 4). Therefore, the receivables with a short remaining term (under one year) can now be determined. Advance payments must be shown directly within fixed assets and inventories (§ 266 Abs. 2).

The task for the person analysing the balance sheet to find hidden reserves in the fixed or current assets is now easier as in accordance with § 252 Abs. 1 Nr. 6 for all legal forms and types of business consistency of valuation is prescribed.[12] Companies become essentially restricted in building up or releasing hidden reserves (compare § 279 ff.). The undervaluation, although still permissible, becomes clear due to classification und information provisions (compare §§ 274, 281, 284 Abs. 2).

3.2 Capital Structure (Equity & Liabilities)

In future, the equity will be disclosed en bloc and classified as to its sources.[13] Both the income retention policies and the level of self-financing become evident.[14] Additionally, an "adjusted" equity can be determined if, for example, omitted pension accruals[15] are being subtracted and undervaluations in the current assets are being added.[16]

The accruals are presented more clearly through subclassification[17] and through the notes to the financial statements (§ 285 Nr. 12). The permitted accruals for expense matching (§ 249 Abs. 2)[18], however, will become a problem as these will provide a variety of possibilities. Some of these, on particular occasions, may have a dubious nature. As the accruals for expense matching are not tax deductible, they can be expected to occur in profitable enterprises.[19]

With regard to the liabilities, a wide range of new additional information is now available. Companies have to note the amount of liabilities with a remaining term of up to one year. The notes to the financial statements must include the total amount of liabilities with a remaining term of more than five years (§ 285 Nr. 1a). Furthermore, it must be mentioned the total amount of liabilities which are secured by mortgages or other rights giving the type and form of security.

Large companies must give this information for each separately shown liability item, and thus the indebtedness of a large company can be clearly discerned in a "liability survey".[20]

Due to the disclosure requirements, five aspects can be recognized:

- maladjusted structure, e.g., too many short term liabilities,
- criteria for the maximum lines of credit or preferred forms of security for certain groups of creditors.
- the total amount of other financial commitments[21] which do not appear in the balance sheet (§ 285 Nr. 3), e.g., invisible financing through leasing,
- liabilities between affiliated enterprises,
- concerning limited liability companies, the liabilities relative to the shareholders (§ 42 Abs. 3 GmbHG).

3.3 Financial Structure (Horizontal Key Figures)

In practice, especially in the area of credit checks through banks, the horizontal key figures are ascertained repeatedly. Due to the new disclosure of the remaining liability terms (up to now original liability term), it is now possible to divide the liabilities into term periods. In particular, the short term debts can clearly be determined.

3.4 Statements of Changes in Financial Position/Cash Flow Analysis

The statement on sources and application of capital, in spite of university professors' recommendations[22], did not become compulsory for the companies. In order to analyse the financial position under the new law, however, the statement of changes in financial position is indispensable.

The cash flow is one of the most important key figures of liquidity. To analyse the cash flow according to the New Accounting Directives Law, the short formula used hitherto[24] has to be modified.[25]

4. The Possibilites of Analysing the Profit and Loss Account

4.0 General Remarks

For sole proprietorships and partnerships there is, apart from the balance sheet, no regulation for a classification format of the profit and loss account (§ 247). This is surely a handicap for the external analyser.

Companies can present the profit and loss account in vertical form following the methods of total costs of sales (§ 275 Abs. 1). Even though each format (§ 275 Abs. 2 and 3) on its own may provide enough information, the option of application will make a branch comparison difficult.

4.1 Analysing the Result

In the area of result analysing, the first concern is the amount of the annual result (net income for the year according to § 275 Abs. 2 and 3) and the changes as compared with the previous year; a serious problem in this connection could be accounting policies affecting the result.[26]

The main problems for the analyst were, up to the present time, the possible choice of valuation methods and the missing consistency concept. The situation has improved considerably because the hidden reserves made permittedly are now more transparent and companies are strongly restricted in building them up. With the relevant information in the notes to the financial statements, it is possible to derive an "adjusted" net income from the shown net income for the year.[27)]

The introduction of the consistency concept with regard to valuation will ensure a better comparison of the annual results in the time series; accounting policies by means of changes in valuation methods are in future difficult.

The profit as per tax balance sheet is the most reliable way to measure the success. A great number of medium size and small enterprises are preparing only one balance sheet anyway, namely the tax balance sheet. With regard to these companies, the estimation of the tax balance profit will become easier since taxes on earnings have to be disclosed. Considering the complex structure of the corporation tax (key-word: equity available for distribution), it is dubious if one could progress from those taxes to a sensible conclusion about the taxable result.

As far as the structure of the result is concerned, it must be stated first that the external analyst has no or only very restricted possibilities to interpret the factors of success of small and medium size companies. Small companies are not obliged to disclose their profit and loss account (§ 326). Medium size companies may combine certain items under the heading "Gross Results" (§ 276).

For the external analysis, the classification format according to § 275 appears at first attractive because it breaks down the annual result into area results, namely

	Operating results
+/-	Financial results
=	Results from ordinary activities
+/-	Extraordinary results
-	Taxes
=	Net income/net loss of the year

According to the definition of the extraordinary results (§ 277 Abs. 4), all items relating to another financial year and arising from ordinary activities are included in the operating result. However, the aperiodical items must be commented on in the notes to the financial statements if they are material (§ 277 Abs. 4 Satz 3). Particularly problematic is the item "Other operating expenses/income" since this includes changes of accruals and of the reserve for an increase in prices as well as special items which have a portion of equity.

The financial result includes income from participations, securities and long term loans, other interests (each of which is marked "of which from affiliated enterprises"), and amortizations and interest expenses.

The items "Extraordinary Income" and "Extraordinary Expense" may only include income and expenses which arise outside the ordinary activities of the company.[28] This could cause great difficulties of classification in some cases.

The area results mentioned are before tax. Certainly it is a great improvement that comments have to be made now in the notes to the financial statements: Comments concerning the extent to which taxes on income and profit affect the results from ordinary activities and the extraordinary results (§ 285 Ziff. 6).

To summarize, the new profit and loss account format and the related disclosure requirements for the notes to the financial statement will especially give the investor a much better view of the profit situation of the company.[29]

5. Statutory Audit Requirements

The following aspects show the new dimensions of the statutory audit on the financial statements and the management report through an auditor (qualified auditor or certified accountant):

a) The number of companies subject to annual statutory audit has increased considerably, even though small stock corporations are now exempt because medium size and large limited liability companies are now also subject to annual statutory audit.

b) The prescribed audit opinion of the auditor, as compared with the audit opinion regulated in the present Stock Corporation Law and the Publicity Law, has been extended:

   - The auditor must state specifically: "The financial statement presents, in compliance with required accounting principles, a true and fair view of the net worth, financial position and results..."[30] (§ 322 Abs. 1). Note however, that the concept of "financial position" is not defined in the law transcripts.[31]

- The audit opinion must eventually be modified "in order to avoid a wrong impression concerning the nature of the audit and the scope of the opinion" (§ 322 Abs. 2 Satz 1).

c) Although the scope of the audit has not changed considerably, the auditor's obligation to report has been extended.[32] "Negative variances compared with the previous year in the net worth, financial position and results and losses which are material to the net income of the year must be included and adequately explained (§ 321 Abs. 1 Satz 4)." The so-called "obligation to declare" for the auditor which, up to now, was a Stock Corporation Law regulation, has been now adopted in the Commercial Code (§ 321 Abs. 2).

d) The registered managers of a limited liability company must submit the audit report to the partners without delay (§ 42a Abs. 1 GmbHG).

6. Disclosure Requirements

All companies (limited liability company, stock corporation and partnership limited by shares) are now obliged to disclose their financial statements (§§ 325 pp.). Small and medium size companies must file their financial statements, notes to the financial statement, management report and, if applicable, the resolution concerning the appropriation of results with the pertinent Commercial Register where anybody has the right to view the documents and to take copies. Details of the Commercial Register in which the documents are filed must be disclosed in the Federal Gazette. Large companies are fully subject to publication in the Federal Gazette. In addition, documents must be filed with the appropriate Commercial Register with the possibility for anyone to examine them.

The Register Court has to check the completeness of the submitted documents and, as far as applicable, their publication.

7. Possible Effects

7.0 Recognizable Problem Areas

The effects of the Accounting Directives Law have not yet been verified through experience. However, there are recognized some problems which can only be illustrated casuistically:

1. As regards the notes to the financial statements, large companies have to break down their sales by areas of activity and by geographically defined markets, if these differ significantly.

Cause for concern: The cartel authorities will gain insight, disclosure of marketing strategies.

2. Small and medium size companies are apprehensive about the disclosure requirements as to the remunerations of their (partner-) managers.

3. The disclosure requirements may go too far. Systematical review of annual financial statements could reveal gaps in the market. Problems could arise for the sub-contractor if the bulk buyer realizes the good profit the sub-contractor makes. Will the bulk buyer become interested in manufacturing himself, or will he try to reduce the price?[33]

4. Due to the standardization of the accounting regulations in the EC, it is not only possible to make intercompany comparisons of financial statements. Also the merging of especially small and medium size companies and the introduction of an EC stock corporation are made easier.

5. The accruals for expense matching now permitted will provide a variety of possibilities. Some of them could be of a dubious nature. On the other side, this makes possible provisions for future expenses which companies must incur.

6. According to prevailing opinion, companies are no more permitted to form hidden reserves at random, however, sole proprietorships and partnerships may still do so. For partnerships, it is also possible to liquidate secretly concealed reserves which could be of danger for not managing partners.

7. The wording of the audit opinion is problematic as far as the net worth, financial position, and results of the company are concerned.

   - The concept "financial position" is not defined but is part of the so-called "true and fair view". According to the prevailing opinion in literature, no changes occurred when compared with the Stock Corporation Law apart from the insertion in § 149 AktG 1965 "within the frame of the valuation requirements".
   - The financial position derived from the financial statement has a static view. Long term engagements, future incoming and outgoing payments as well as investment plans are not being considered, not even in the notes to the financial statement or in the management report.

- Auditors are worried that the new wording of the audit opinion could be misleading.[34)]

8. The operational format in the profit and loss account is widely used internationally but German firms will have, at least in the beginning, difficulties if used with regard to competitors (e.g. pharmaceutical industry). The operational format uses another manufacturing cost concept (§ 275 Abs. 3 Ziff. 2) as it is defined for valuation purposes (§ 255 Abs. 2) in the financial statement.

9. Concerning the optional requirement to accrue in full pension obligations, there are partly doubts as far as the disclosure requirements of the deficit in the notes to the financial statement are concerned. On the other hand, there are no regulations as to how to arrive at it (method, interest rate).[35)]

10. Banks have a positive view of the new accounting law as far as the rating of their customers is concerned;[36)] the reasons are:

    - The financial statements must be prepared within prescribed terms.
    - Due to the new classification and valuation requirements, new findings are possible.
    - In future, important information has to be given in the notes to the financial statement.
    - Due to the classification requirements, intercompany comparisons of financial statements become easier.
    - The amount, development, and deficit of equity are easier to recognize.
    - The amount of the separately shown revenue reserves gives a conclusion of the profit position and retained income policy.
    - Receivables from and liabilities payable to affiliated enterprises and participations as well as against partners of limited liability companies must be shown or noted separately.
    - The kind of financing is made more transparent.
    - The profit and loss account shows more business data as now the operating, financial and extraordinary results have to be disclosed.
    - The valuation requirements are now mostly coded and therefore fixed.
    - The management report must contain a description of the development of the business, the anticipated development of the company and post balance sheet data events of special importance.

    The above aspects are also, more or less, valid for investors of equity.

7.1 Experiences to Date

Presently, qualified auditors, tax advisors and solicitors are, in their advisory activities, mainly confronted with questions on how to avoid or minimize the disclosure requirements.[37] The following ideas should be mentioned:

1. Dislocation of certain activities into a partnership whereby the partnership receives adequate payment in form of a tenancy, consultancy or employment agreement
   a) Dislocation of areas such as:
      - development and research in exchange for licence agreements or know-how,
      - quality control, marketing, sales promotion,
      - movable fixed assets - sale and lease back.
   b) Dislocation of earnings:
      - sales via own partnership.
   c) Dislocation of personnel into a partnership and lending of said personnel to the company whereby old age pension obligations stay with the company.
2. Change of seat and name of the business to avoid disclosure.
3. Dislocation of foreign business into foreign enterprises not in the EC, mainly to USA or Switzerland.
4. Reduction of balance sheet result which must be disclosed by making accruals which are not tax deductible.
5. Change of leasing agreements so that the company is liable for maintenance of substance and, therefore, has to make accruals.

8. Conclusion

An attempt is made to show the Principal-Agents-Information Structure by means of selected areas and with the help of specific examples. Many of the aspects referred to are valid for either investors (principals) or agents (managers). Other aspects, however, are valid for both investors and managers, even though the importance may differ according to the respective interests.

Footnotes

1) dated 12/19/1985, Bundesgesetzblatt I 1985, S. 2355; this refers to a pure modification law which revised especially the Commercial Law. If not mentioned otherwise, all cited regulations concern the amended Commercial Law

2) compare to this Spremann (1985, pp. 34)

3) supplementary compare Wöhe (1986, pp. 908)

4) compare Blumers (1986, p. 2033)

5) Geuer (1982, p. 342)

6) Göllert (1984, p. 1845)

7) Göllert (1984, p. 1853), Biener (1979, p. 1, p. 14)

8) for further particulars and method to show the fixed assets gross observe Göllert (1984, p. 1845)

9) Göllert (1984, p. 1845)

10) Zilias (1986, p. 1110)

11) for further particulars observe Küting (1985, p. 1089)

12) compare for many Pfleger (1986, p. 1133)

13) compare Harms/Küting (1983, p. 1449), Göllert (1984, p. 1847)

14) Göllert (1984, p. 1847)

15) Heubeck (1986, p. 317, p. 356), Höfer/Lemitz (1986, p. 426)

16) for particulars observe Göllert (1984, p. 1848)

17) compare Göllert (1984, p. 1848), Küting (1985, p. 1096)

18) particulars and further literature references gives Siegel (1986, p. 841)

19) regarding the problems of latent taxes which also arise in this area observe amongst others Siegel (1984, p. 1909), Harms/Küting (1985, p. 94), Heydkamp (1986, p. 1345), Schneeloch (1986, p. 517), Kugel/Müller (1986, p. 210), Weyand (1986, p. 1135)

20) for further particulars and presentation of the liability survey observe Hoffmann (1983, p. 10), Göllert (1984, p. 1850)

21) to this compare the comment of the committee "Rechnungslegungsvorschriften der EG-Kommission" der Gesellschaft für Finanzwirtschaft in der Unternehmungsführung eV. (GEFIU), Thesen zu ausgewählten Problemen bei der Anwendung des BiRiLiG, Der Betrieb 1986, p. 1985, here p. 1986

22) Die Betriebswirtschaft 1979, 4 (These 7) and especially p. 30

23) to this compare Schoenfeld (1985, p. 561)

24) compare Coenenberg (1982, p. 375)

25) particulars gives Göllert (1984, p. 1850)

26) to this compare Wöhe (1985, p. 715, p. 754)

27) particulars gives Göllert (1984, p. 1852)

28) to this compare Leffson (1986, p. 433), Niehus (1986, p. 1293)

29) at the same conclusion arrives Geuer (1982, especially p. 344)

30) to this compare Ballwieser (1985, p. 1034)

31) to this compare Sonderausschuß BiRiLiG des Instituts der Wirtschaftsprüfer, zur Darstellung der Finanzlage im Sinne von § 264 Abs. 2 HGB, Die Wirtschaftsprüfung (1986, p. 393)

32) to this compare Emmerich/Künnemann (1986, p. 145), Ludewig (1986, p. 377)

33) There are also doubts under constitutional law arising from Art.3/12 Grundgesetz , since now all limited liability companies are liable to the disclosure requirement; the doubts were impressively described by Friauf (1985, p. 245)

34) to this compare Gmelin (1986, p. 60)

35) particulars gives Heubeck (1986, pp. 325)

36) to this compare Geuer (1982)

37) even in the specialized literature, considerations of this kind are being (openly) made. Compare for instance Tillmann (1986, p. 1319), Woltmann (1986, p. 1861).

References

Ballwieser, W. (1985): Sind mit der neuen Generalklausel zur Rechnungslegung auch neue Prüfungspflichten verbunden?, Der Betriebs-Berater, Heft 16, 1034-1044

Biener, H. (1979): Auswirkungen der Vierten Richtlinie der EG auf den Informationsgehalt der Rechnungslegung deutscher Unternehmen, Betriebswirtschaftliche Forschung und Praxis, Heft 1, 1-16

Blumers, W. (1986): Neue handels- und steuerrechtliche Bilanzierungsfristen und die Risiken der neuen Rechtslage, Der Betrieb, Heft 40, 2033-2036

Coenenberg, A.-G. (1982): Jahresabschluß und Jahresabschlußanalyse, 6th ed., München

Emmerich, G.; Künnemann, M. (1986): Zum Lagebericht der Kapitalgesellschaft, Die Wirtschaftsprüfung, Heft 6, 145-152

Friauf, K. (1985): Die Publizitätspflicht für Gesellschaften mit beschränkter Haftung aus verfassungsrechtlicher Sicht, GmbH-Rundschau, Heft 8, 245-253

Geuer, H. (1982): Auswirkungen des Bilanzrichtlinien-Gesetzes auf die Kreditinstitute als Gläubiger, Die Wirtschaftsprüfung, Heft 12, 340-344

Gmelin, H.J. (1986): Neue Anforderungen an die Darstellung des Prüfungsergebnisses zum Einzel- und Konzernabschluß, in: Bericht über die Fachtagung 1986 des Instituts der Wirtschaftsprüfer e.V., Düsseldorf

Göllert, K. (1984): Auswirkungen des Bilanzrichtlinien-Gesetzes auf die Bilanzanalyse, Der Betriebsberater, Heft 29, 1845-1853

Harms, J.; Küting, K. (1983) Ermittlung und Ausweis des Eigenkapitals nach dem Bilanzrichtlinien-Gesetz, Der Betrieb, Heft 28, 1449-1454

Harms, J.; Küting, K. (1985): Probleme latenter Steuern im Entwurf des Bilanzrichtlinien-Gesetzes, Der Betriebs-Berater, Heft 2, 94-101

Heubeck, K. (1986): Betriebliche Versorgungsverpflichtungen nach dem neuen Bilanzrecht I,II, Die Wirtschaftsprüfung, Heft 12, 317-328, Heft 13, 356-364

Heydkamp, W. (1986): Grundsätzliche Überlegungen zur Steuerabgrenzung im Einzelabschluß nach dem neuen Bilanzrecht, Der Betrieb, Heft 26, 1345-1347

Höfer, R.; Lemitz, H.-G. (1986): Betriebliche Altersversorgung und das neue Bilanzrecht, Der Betriebs-Berater, Heft 7, 426-433

Hoffmann, W.-D. (1983): Praxisorientierte Einführung in die Rechnungslegungsvorschriften des Regierungsentwurfs zum Bilanzrichtlinien-Gesetz, Der Betriebs-Berater, Beilage 1 zu Heft 5, 10-21

Küting, K. (1985): Externe Liquiditätsanalyse auf der Grundlage der Bilanz nach künftigen Bilanzrecht, Der Betrieb, Heft 21, 1089-1098

Kugel, B.; Müller, U. (1986): Die Bilanzierung latenter Steuern nach § 274 HGB, Die Wirtschaftsprüfung, Heft 8, 210-217

Leffson, U. (1986): Der Ausweis des Außerordentlichen nach dem HGB, Die Wirtschaftsprüfung, Heft 16, 433-436

Ludewig, R. (1986): Der Bestätigungsvermerk gemäß § 322 HGB im Hinblick auf den Lagebericht, Die Wirtschaftsprüfung, Heft 14, 377-380

Niehus, R. (1986): Aufwendungen und Erträge aus der "nicht gewöhnlichen Geschäftstätigkeit" der Kapitalgesellschaft: Abgrenzungsfragen zum Ausweis der außerordentlichen Posten nach neuem Recht, Der Betrieb, Heft 25, 1293-1297

Pfleger, G. (1986): In welchen Ausnahmefällen darf vom Grundsatz der Bewertungsstetigkeit abgewichen werden?, Der Betrieb, Heft 22, 1133-1136

Schneeloch, D. (1986): Probleme der Steuerabgrenzung im Einzelabschluß, Die Wirtschaftsprüfung, Heft 19, 517-528

Schoenfeld, A. (1985): Die Kapitalflußrechnung als Bestandteil eines Finanzlagenpostulats nach geltendem und zukünftigem Bilanzrecht, Der Betriebs-Berater, Heft 9, 561-565

Siegel, Th. (1984): Probleme latenter Steuern im Entwurf des Bilanzrichtlinien-Gesetzes, Der Betriebs-Berater, Heft 30, 1909-1916

Siegel, Th. (1986): Echte Aufwandrückstellungen und der Wandel des Gesellschaftsschutzes im neuen Bilanzrecht, Der Betriebs-Berater, Heft 13, 841-844

Spremann, K. (1985): Finanzierung, München

Tillmann, B. (1986): Umwandlung auf doppelstöckige GmbH & Co KG - Ein Ausweg aus der Publizitätspflicht der GmbH?, Der Betrieb, Heft 25, 1319-1323

Weyand, S. (1986): Die Bilanzierung latenter Steuern nach § 274 HGB, Der Betrieb, Heft 23, 1185-1189

Wöhe, G. (1985): Möglichkeiten und Grenzen der Bilanzpolitik im geltenden und im neuen Bilanzrecht, Deutsches Steuerrecht, Heft 23, 715-721, 754-761

Wöhe, G. (1986): Einführung in die Allgemeine Betriebswirtschaftslehre, 16th ed., München

Woltmann, A. (1986): Bilanzrichtlinien-Gesetz und Gesellschaftsrecht Statutenänderungen und Unternehmensumgestaltungen, Der Betrieb, Heft 36, 1861-1866

Zilias, M. (1986): Zum Unternehmensbegriff im neuen Bilanzrecht (Drittes Buch des HGB), Der Betrieb, Heft 21, 1110-1112.

Auditing in an Agency Setting

Wolfgang Ballwieser

Summary: Agency theory is concerned with contracts which lead to optimal incentives and risk-sharing. The purpose of this paper is to examine the influence of auditing on both aspects, when audits are performed (i) by an owner of a firm (who is a principal) and (ii) by an auditor in order to motivate the firm's manager (who are both agents). We especially ask, under what conditions the owner can expect truthful financial reporting from the manager and a truthful report by the auditor. We further ask, whether it is likely to expect coalitions of the manager and the auditor against the owner. Since all results are gained in one-period agency models with at most two agents, the stability of results and the practical relevance of the models are discussed. Though the models, up to now, have no decision-supporting function they tell us that coalition-forming of agents against the principal seems to be likely if there are no other factors which are neglected in agency theory so far.

Table of Contents:

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

1. The Problem

An agent, e. g. a manager, who is supposed to perform his service on behalf of a principal, e. g. an owner of a firm, will always be obliged to give some information about the results of his efforts in order to allow for the principal to check whether his aims are being realized, or not. The reporting would only be unnecessary if (i) the interests of both the principal and the agent were identical and if (ii) the competence of the manager were at least as great as the competence of the owner. Although the second condition is fulfilled in general, the identity of interests cannot be expected in most of the economic agency relationships. If the agent is allowed to select the rules of reporting, concerning, for example, timing, content or reliability, autonomously, then the principal might have reason to be suspicious of not being informed properly. This is one of the reasons why the commercial law usually restricts the "coarseness" of the financial reporting of those firms who must publish their balance sheets and profit and loss statements by means of GAAP (Generally Accepted Accounting Principles) and further detailed regulations (cf. Ng 1978; for strong conditions to leave some discretion for reporting to the manager cf. Demski, Patell and Wolfson 1984; Ballwieser 1985, pp. 35-36, and Verrecchia 1986).

Since the owners cannot judge whether the information of the management corresponds to these rules, the balance sheets and profit and loss statements of those firms generally have to be audited, and the auditor's opinion about the reliability of the information has to be published. (In Germany, there is an exception of obligatory auditing only for so-called small firms, as sections 267 and 316 Handelsgesetzbuch (HGB) show).

An audit performed by a third person is not necessarily the only means for helping the owner of a firm. Perhaps the owner could manage the audit by himself. The reasons for engaging an auditor might be due to the lacking competence of the equity-holder or simply the physical distance of the owner from the manager (cf. ASOBAC 1972, p. 26). But since competence may be gained in the long run, and distance can be reduced even in the short run, the decision to employ an auditor should have economic reasons, at least if we neglect obligatory audits by independent auditors because of the restrictions of law. Those rules cannot always be explained as the result of a cost-benefit-analysis of a single owner of a single firm.

Though an auditor may be very valuable to an owner, it would be rather myopic for him not to be aware of the problems which the auditor can also create. Why should it be obvious that he will act on behalf of the owner if the interests of both parties are not identical? Clearly, the identity of interests will only be given by chance. Therefore, the auditor must be seen as another agent who should also inform the principal about the results of his efforts, so that the principal can control whether his aims are being realized, or not. This, in practice, leads to the auditor's report which must be given to the management and the board of directors.

This paper is dedicated to the problems that arise in addition to the well-known incentive and risk-sharing problems of a two-person agency relationship between owner and manager (cf. Rees 1985) when an auditor is engaged in order to support the owner in influencing his manager. To be a bit more precise, the paper is concerned with the questions of what the (sufficient) conditions are that make (i) an audit and (ii) an auditor valuable to an equity-holder. We are especially interested in the questions of: (i) under what conditions the owner can expect unbiased ("truthful") financial reporting by the manager, (ii) under what conditions the owner can expect unbiased reporting by the auditor and (iii) whether it is likely that a coalition of the manager and the auditor against the owner will be formed.

The answers to those questions shall be given by means of formalized agency theory (cf. Arrow 1985, Namazi 1985, Baiman 1982 for an overview of the state of the art). The models which are part of that theory will not be elaborated in detail in this paper because of their complexity and length. Some important results which can be gained with their help, however, will be described in a systematic manner.

The paper is organized as follows. Firstly, the basic premises of agency theory models will be shortly described. We sketch a two-person one-period model at the beginning, since this simple model allows us to discuss the optimality of auditing by the owner himself. Then this simple model will be extended by the auditor as a third person. After discussing the most important results of three-person one-period models, the paper culminates with observations on the relevance of the results (and therefore the relevance of the underlying models) for practical considerations.

2. A Basic Model

The following model may serve as a starting point for our considerations:

The owner of a firm and a manager are considering, whether they should form a contract for an agency relationship, or not. Within that relationship only the manager takes productive actions on some resources of the owner which affect the welfare of both parties. The contract specifies a sharing rule, a rule which determines a priori what part of the (random) financial results of the productive actions chosen by the manager will be transferred from the manager to the owner. We call the financial result "cash flow".

Both the owner (who is a principal) and the manager (who is an agent) behave as if they maximize the expected value of von Neumann - Morgenstern utilities. The only argument of the owner's utility function is the cash he receives. He is supposed to be (weakly) risk averse. The arguments of the utility function of the manager are cash and effort. The manager is risk averse, too. Furthermore, he suffers disutility of effort. This disutility will be assumed separable from the utility of cash.

The effort of the manager which is representative of the productive action that he chooses affects the cash flow of the firm, together with the occurrence of an stochastic event (a state of the world) that cannot be influenced by the manager. In other words, the cash flow is a random variable whose distribution depends on the action taken by the manager. The distribution of the cash flow to a higher effort stochastically dominates that to a lower one; that is, effort increases the likelihood of a favorable cash flow. That makes the effort of the manager valuable to the owner; on the other hand, he must consider that there is more disutility experienced by the manager in increasing effort.

Unfortunately, the owner cannot observe the manager's realized effort, nor the realization of the stochastic event. Because of this the owner cannot observe the realized (true) cash flow even if he knows precisely the function specifying cash flow as a variable of the effort of the manager and random event. Both is only observable by the manager, and he reports about the cash flow. His fee functionally depends on his report. If there is no other information that the owner can gain, the reported cash flow will be the only argument of the sharing rule. In order to distinguish different models which will be of interest later on, we call the type of model with nothing more than reporting information model 1. The manager pays the part of reported cash flow which has been agreed upon in the contract to the owner. He receives the difference of realized cash flow and the amount paid to the owner. (This direction of the transfer is only due to the fact that the owner cannot observe the realized cash flow.) Model 2 is different from the formerly described type in that the owner can gain some further information, i. e. a signal that is, in general, different from reported cash flow, but not stochastically independent from the true (realized) cash flow. This signal may be a tax amount or a wage claim.

Although the owner cannot actually observe the realized cash flow, he has perfect knowledge of the preferences of his manager. He also shares exactly the manager's probability expectations of the state of the world that affects cash flow, together with the manager's effort. He even knows the cash flow function. The owner anticipates that the manager may misrepresent realized cash flow in his report. Given a useful technology of auditing, this is the reason why he may be successful in giving incentives to the manager so that he will act on behalf of the owner even though he is maximizing his own utility. To make the problem complete, the owner has to consider that the manager bears opportunity losses when he is employed by the owner. Therefore, he has to guarantee the manager an expected utility that is at least as high as the highest opportunity loss he deserves.

Solutions to the problem of contracting must have the properties of Nash-equilibria. That means in the given context that both parties must find it in their own best interest to form the contract. When there are different equilibria the optimal ones must be Pareto-efficient.

In order to specify the sequence of events when there are optimal contracts to both parties the following diagram for model 1 may be helpful (cf. Antle 1981, p. 26):

time →

Owner chooses sharing rule $s(\hat{x}(x))$.	Manager chooses effort e and reporting system (reporting function) $\hat{x}(x)$.	State of the world Θ and cash flow x realize.	Cash flow $\hat{x}(x)$ is reported and amount $s(\hat{x}(x))$ is paid to the owner. Manager receives x-s(.).

Fig. 1: Sequence of events in model 1

The symbols mean:

e	effort (or action) chosen by the manager,
Θ	realized state of the world,
x	realized cash flow, dependent on effort and state of the world,
$\hat{x}(x)$	reported cash flow, dependent on realized cash flow,
$s(\hat{x}(x))$	sharing rule, dependent on reported cash flow.

For model 2 the sequence of events will be described by:

time →

Owner chooses sharing rule $s(\hat{x}(x),y)$.	Manager chooses effort e and reporting system $\hat{x}(x)$.	State of the world Θ, cash flow x and signal y realize.	Cash flow $\hat{x}(x)$ is reported and $s(\hat{x}(x),y)$ is paid to the owner. Manager receives x-s(.).

Fig. 2: Sequence of events in model 2

In addition to Fig. 1 there is one new symbol which also alters the arguments of the sharing rule:

y	signal which is not stochastically independent from realized cash flow x,
$s(\hat{x}(x),y)$	sharing rule, dependent on reported cash flow and signal.

3. The Value of Auditing [1]

Founded on the basic model there are four approaches which discuss the value of auditing: Ng and Stoeckenius (1979), Woodland (1981), Yandell (1981) and - in a somewhat different way - Penno (1985). There is, of course, more literature on monitoring, e. g. Harris and Raviv (1978) and (1979), Shavell (1979) and Holmström (1979), but monitoring may be distinguished from auditing in the formal sense that auditing means verification of an agent's report whereas monitoring is information-gathering about the effort or the state of the world both of which influence financial results (cf. Baiman 1979, p. 25; "monitoring" is used in a different sense by Demski, Patell, and Wolfson 1984, p. 16). Based on that definition the value of auditing is discussed in the four models without the existence of an real auditor, who creates additional incentive and risk-sharing problems beyond those of the owner-manager relationship. Even if the audit is called to be performed by an auditor (as is the case in the Ng and Stoeckenius model, e. g.), there is no person who will create agency problems. Therefore, the audit really could be done only by the owner himself.

With auditing, the diagrams of the last section will be modified to:

time →

Owner chooses sharing rule s(.) and decides on audit level and penalty function for detected errors in reporting.	Manager chooses effort e and reporting system $\hat{x}(x)$.	State of the world θ , cash flow x (and signal y) realize.	Cash flow $\hat{x}(x)$ is reported.	Audit result made known to the owner and the manager.	Manager transfers cash dependent on s(.) and audit result.

Fig. 3: Sequence of events with audit activity by the owner

1 For some parts of the discussion in sections 3 and 5 also compare Ballwieser (1987). More details can be found here.

Ng and Stoeckenius develop a model where the fee of the manager increases in reported cash flow. The owner has an audit technology with constant costs for each unit of quantifiable audit level. The purpose of the audit is to form an opinion as to whether the manager's report is in error. The probability of detection that the report is in error increases with respect to the error's size (which is the difference of reported and true cash flow) and audit level. The manager will be penalized if an error is detected, and the penalty function is based on the size of the error.

The authors show, that costless auditing is valuable to the owner, since a truthful (unbiased) reporting of the manager can therefore be induced. Of course, it has been also shown that unbiased (truthful) reporting may be gained without auditing (cf. the so-called revelation principle in Myerson 1979, also compare Levinthal 1985, pp. 66 - 73 or Rees 1985, p. 86). But in order to reach that result it seems to be necessary that the manager's remuneration is paid independently of realized cash flow. Consequently, there is no incentive for him to make a great effort which imposes a cost (in form of an opportunity loss) to the owner (cf. Antle 1982, p. 512). This is different from the model of Ng and Stoeckenius, since, for proving the result, they have assumed that the fee for the manager follows a concave function of reported cash flow. Further, it has been assumed that the expected amount of remuneration is less than the expected amount of the firm's cash flow.

Ng and Stoeckenius further prove, that costly auditing is valuable to the owner when he is risk-neutral and further conditions are fulfilled. "Specifically, under appropriate conditions, the demand for audit effort level can be shown to be downward sloping with respect to its price." (Ng and Stoeckenius 1979, p. 15)

What seems to be remarkable at first is the inducing of unbiased reporting of the manager by means of auditing. The crucial assumptions are, of course, that auditing is costless and the audit technology is perfect in the sense that the detection of an error leads to an exact knowledge of the error's size. Therefore, the unbiased reporting is intuitive (cf. Ng and Stoeckenius 1979, p. 12). The penalty can always be set high enough to induce the manager to report truthfully. This is also the case when the exactness of the audit technology is slightly weakened as it has been suggested by Baiman (1979, p. 28).

It is hardly possible to discuss the exact properties of the audit technology, because there is no detailed description thereof. In addition, it is notable that optimal fee schedules (derived from a two-person model) may easily violate one or both of the two assumptions about the properties of the fee function (concerning concavity and expected amount of remuneration) which have been used to gain the first result. Counterexamples may be found in Holmström (1979, p. 79) and Woodland (1981, p. 41). Therefore, the assumptions are rather restrictive, though output-independent remuneration could be avoided.

A good description of audit technology that is comparable with usual assumptions in textbooks on statistical auditing (e. g. Roberts 1978; Bailey 1981) is provided by Woodland. The audit level or audit intensity which he describes reflects sample size in a statistical sampling procedure. The result of auditing (a signal y) is an unbiased and consistent estimator of realized cash flow x. The owner must keep in mind that the manager may overreport or underreport and must decide on the audit level in view of the α and β risk factors as they are well known from testing hypotheses (cf. Roberts 1978, pp. 40-48; Bailey 1981, pp. 57-83). The manager "will be penalized if the audit result differs, by some predetermined amount, from the financial report". (Woodland 1981, p. 57) The owner has to decide on tolerable α and β errors which affect the probability that the manager will be penalized. He also has to decide on the penalty function.

Woodland shows that in the case of overreporting by the manager, "in general, it cannot be concluded that a contract based on output will strongly dominate a flat-wage contract." (Woodland 1981, p. 66) But the strong domination is one of the necessary conditions for gains from auditing (together with the unobservability of cash flow and the manager's risk aversion). He further proves that under certain premises which cannot be easily explained economically, there exists a finite level of audit intensity (sample size) that is truth-inducing. Then, the manager will report the realized cash flow without any error (cf. Woodland 1981, p. 86). If the auditing technology is truth-inducing, the noisy signal of cash flow is as valuable as perfect knowledge of cash flow. That means that the owner can construct a contract that is equivalent to the optimal realized (!) cash flow based contract (cf. Woodland 1981, p. 113). Taking into consideration that the manager may over- or underreport, it seems plausible for Woodland that the demand for auditing will not be a strictly decreasing function with respect to the price of auditing services, but a step function. This is due to the fact that price variations influence the optimality of output based contracts with auditing versus flat-wage contracts without auditing. Under certain circumstances, for example, the principal would either demand no auditing services (if the price is "too high") or only that level of audit effort that is truth-inducing (if the price is "low enough"; cf. Woodland 1981, pp. 98-100).

Somewhat different from the formerly discussed models are the models of Yandell (1981) and Penno (1985). Yandell assumes - in contrast to Ng/Stoeckenius and Woodland - that the owner suffers a degree of disutility of effort, also. He therefore develops a hierarchical model with "society" as a second principal trying to motivate the owner. Since his results are heavily based on examples which cannot be easily generalized, we chose to neglect his model at this point.

Penno shows that it is to the advantage of the manager to "voluntarily" give a financial report in the case that the owner gains imperfect verification of cash flow x by means of costless auditing. The added communication leads to a contract with better risk-sharing and a strict Pareto-improvement if the manager's report comes earlier than the signal produced by the audit technology. The result is intuitive, because the imperfect verification of cash flow x by means of auditing imposes risk on the compensation of the manager which can be reduced by means of own reporting (cf. Penno 1985, pp. 245-246). Communication allows basing the sharing rule on $\hat{x}$ and y instead of y alone, where y is the audit technology's signal. The difference between Penno's model and the models, discussed formerly, is that Penno starts with the assumption that there would be auditing without financial reporting. He, therefore, is looking for the gains of financial reporting instead of those of auditing. But to be consistent with our definition of auditing given at the beginning of this section, we have to assume that financial reporting is the starting point. Auditing may lead to better efforts and better reporting on the part of the manager.

Summing up the results, it has been shown that auditing may encourage a manager to report truthfully. The truthful reporting may be gained without an output-independent remuneration of the manager which would not help to mitigate the problem of incentive in an agency relationship. The demand function for auditing services may not strictly decrease with respect to price. The audit technology can be understood as a sampling procedure with such favorable statistical properties as unbiasedness and consistency.

4. An Extended Model

The basic Model of section 2 will now be extended by the auditor who creates new incentive and risk-sharing problems. The owner now offers two different contracts; one to the manager and another one to the auditor. There are two sharing rules s(.) and r(.) depending on the manager's report $\hat{x}$, the auditor's report $\hat{z}$ and perhaps a signal y, which is not stochastically independent of x and/or z. The auditor's report is a function of his knowledge z which depends on his effort a. The auditor is risk averse and has disutility of effort. He only accepts the contract if his expected utility is at least as high as the expected utility of his best alternative which is exogenously given. The owner knows perfectly the utility function of the auditor, but cannot observe the auditor's realized effort and his true knowledge z. The owner may be risk averse or risk neutral.

In order to discuss the following results we must pay attention to the sequence of the events of effort chosen by manager and auditor, financial reporting, auditor's reporting and signal's gaining. If manager and auditor do not work simultaneously, we can demonstrate the sequence of events with the following diagram, which is very similar to the one developed by Noel (1981, p. 104):

Owner chooses sharing rules $s(\hat{x},\hat{z},y)$ and $r(\hat{x},\hat{z},y)$.	Manager chooses effort e and reporting system $\hat{x}(x)$.	State Θ and cash flow x occur.	Manager reports cash flow $\hat{x}(x)$.	Auditor chooses effort a and reporting system $\hat{z}(z)$.	State t occurs and auditor observes z.

time →

Auditor reports $\hat{z}(z)$.	Signal y is generated.	Manager transfers cash dependent on s(.) to the owner. Owner transfers cash dependent on r(.) to the auditor. Manager receives x-s(.).

Fig. 4: Sequence of events with audit activity by the auditor

The new or slightly changed symbols (compared to Fig. 1 and 2) are:

a	effort (or action) chosen by the auditor,
t	state variable affecting audit evidence,
z	audit evidence, depending on realized cash flow x, auditor's effort a and state variable t,
$\hat{z}(z)$	auditor's report, depending on audit evidence,
y	signal which is not stochastically independent from realized cash flow x and audit evidence z,
$r(\hat{x},\hat{z},y)$	sharing rule for owner and auditor, dependent on reported cash flow $\hat{x}$, auditor's report $\hat{z}$ and signal y,
$s(\hat{x},\hat{z},y)$	sharing rule for owner and manager, dependent on reported cash flow $\hat{x}$, auditor's report $\hat{z}$ and signal y.

5. The Advantages and Problems of the Auditor [1]

5.1 Gains from Auditing by an Auditor

There are three models, developed by Noel (1981), Antle (1981) and (1982) and Baiman, Evans III and Noel (1985), which give us some insight into the advantages and the problems caused by the auditor.

Noel shows that auditing is without value if the compensation of the manager and the auditor depends only on their own reports without the owner's opportunity to detect and penalize false reporting (cf. Noel 1981, pp. 94-95; such a model would be a straightforward generalization of model 1 in section 2). Phrased in another way, the signal y is necessary for a positive value of auditing to the owner. Because of that, Noel assumes later on that the owner has access to imperfect information about the cash flow. With this assumption, it can be shown that there exist optimal contracts which induce truthful reports of the auditor. If the auditor has to decide whether to do a perfect

---

1 Also compare footnote 1 at the beginning of section 3.

costless audit (which leads to knowledge of cash flow x with certainty, but presupposes great effort) or no audit at all (which leads to no knowledge, but requires no effort, either), the optimal solution can be shown to always include a perfect audit. "Finally, the optimal contract with financial reporting by the manager can be proved inferior to a feasible auditing contract involving (perfect) audit ..." (1981, p. 129). This means that there is always a positive level of auditing.

A similar result has been derived in the model of Baiman, Evans III and Noel. The authors assume that the owner is risk neutral in order to eliminate that the auditor is only hired to share risk. Then the value of auditing can be traced back to the auditor's influence on financial reporting of the manager. Baiman et al. demonstrate that the auditor will only be engaged when he can be motivated to do an perfect audit. Given perfect audits, the relevant contracts are only those which induce truthful reporting by the auditor. The optimal compensation of the auditor can be shown to be independent of cash flow x and the auditor's report $\hat{z}$. Costs of auditing are considered.

5.2 Problems Caused by the Auditor

One crucial assumption of the formerly discussed models is the perfectness of auditing. But even more important is the negligence of the possibility of coalition-forming on the part of the auditor and the manager against the owner. The coalition-forming cannot be ruled out, but must be taken into account as Antle shows.

Antle develops a somewhat different model compared with the model of Noel, since there are parallel actions of the manager (choosing his effort e) and the auditor (choosing his effort a). The financial report, the auditor's report and the signal y come simultaneously. If we neglect those details, though they are relevant for the optimization calculus, and the reasons given for this sequence of events, we can concentrate our attention on the basic results. Antle shows that randomized strategies instead of pure ones can be optimal within the games of the owner, manager, and auditor. This is in contrast to the basic model (including manager and owner alone) where randomized strategies may be neglected. Antle nevertheless studies only pure strategies, since their handling is much easier.

Further, he shows that it is "not guaranteed that the subgame equilibrium that maximizes the owner's expected utility also maximizes the auditor's and manager's expected utilities. Therefore, there could be a dominant equilibrium which induced subgame equilibria which did not maximize the owner's expected utility over all the equilibria of the subgame". (1981, p. 49) Those dominant equilibria have not been studied, though they might be interesting. Combined with the preceding point it has to be considered that a solution to the owner's contracting problem could "call for the auditor and manager to play an inferior (for them) subgame equilibrium". (1981, p. 50; 1982, p. 520) Since this is no reasonable behavior, Antle looks for some "collective rationality" constraints to avoid it (1981, p. 53 and 1982, p. 521), but those constraints cannot be formulated to solve the (mathematical) problem.

The formulation of the problem by Antle means that he, perhaps, is studying dominated equilibria (because of the restriction on pure strategies) and that some of the equilibria which he derives may not be reasonable. In the case of unreasonable equilibria, coalition-forming on the part of the auditor and the manager has to be expected. If we neglect the fact that randomized strategies may be optimal, it can be shown that within the restricted set of pure strategies, truthful reporting of the auditor can be induced and is optimal behavior.

Summing up the results it has been shown that the auditor may help the owner in the case of having a perfect audit technology, since he can be motivated to do a perfect audit and to report truthfully. But up till now this result is based on models where perhaps dominated pure strategies of the game players have been assumed and a coalition-forming by the auditor and the manager against the owner has been neglected. This last assumption seems to be questionable, at least if there are no convincing reasons why the coalition should not be very probable.

6. Conclusions

6.1 The Stability of Results

The models should be seen as simple, though mathematically somewhat complicated devices to explain the demand for auditing. They are simple since they abstract from multiperiod optimization behavior, from markets which allow some competition between different principals and agents and from institutions which may help to gain trust in contracting partners. They are even somewhat artificial in assuming that the principal has perfect knowledge about such important factors as the utility functions, the probability expectations, and even the technology of the agents, but is not able to use the technology and to observe the results of the agents' actions. Although it must be admitted that the assumptions about the principal's knowledge are not very reasonable, if we expect that they conform to reality, they facilitate modeling. Less restrictive assumptions about the principal's knowledge are welcome, but, up till now, they lead to difficulties of mathematical tractability.

More important seems to be the omission of multiperiod considerations and of markets. Agency theory is concerned with strongly pursued individualistic behavior where each contracting partner tries to maximize his own utility. Since the models which have been discussed have a planning horizon of one period, the problem of cheating and of giving incentives in order to avoid such a behavior may be overstated. Cheating may only be optimal for an agent as one-period behavior if he neglects all further periods where he must try to form other contracts. Of course, our argument implicitly assumes that in a multiperiod planning horizon it is to be expected with some material probability that cheating may be detected and a suitable penalty would be threatening. It further assumes that the principal is not fully dependent on the agent, because this agent is the only person who can do the job. Rather, there must be some form of competition. Competition could mitigate the cheating problem (cf. especially Fama 1980, pp. 292-295).

In creating the models, these problems of course have been foreseen. There are some approaches which are concerned with multiperiod considerations (cf. Fellingham/Newman/Suh 1985, Rogerson 1985, Lambert 1983, Townsend 1982, Radner 1981). For example, Radner showed that in longlasting relationships of principal and agent, the problem of cheating may be alleviated. He provides conditions under which so-called approximate noncooperative equilibria (epsilon equilibria) of the entire sequential game of two players can produce cooperative outcomes of the component subgames. But Radner's model is not compatible with our basic model, since he assumes that the agent first observes a random environmental variable and then chooses his action whose outcome is observable by the principal. According to the description of different types of agency theory models by Arrow (1985, pp. 38-42), the model is of the hidden information type instead of the hidden action type which has been considered here. Radner's assumption that the same one-period situation is repeated a finite number of times, is rather artificial. Furthermore, his model does not allow for the two agents which we have been interested in.

Lambert shows in his multiperiod model that in an optimal long-term contract the agent's compensation in one period does not only depend on his current performance but also on his past ones. If the agent has no option to leave the firm in any prior period before the long-term contract is finished, he acquires the intuitive result "that the more periods the agency relationship lasts, the more the incentive problem is alleviated." (Lambert 1983, p. 448) His model is still of the two-person type with observable cash flow for the owner and the manager. It is unrealistic in the sense that the production functions are separable over time (that means, effort in one period has no effect on cash flow in any other period) and the states of the world are independently distributed over time.

Since it has been shown (for example by Lambert 1983 and Rogerson 1985) that in optimal long-term contracts "memory" plays a very strong role, the qualitative conclusions of one-period models are suspect in a multiperiod setting. But memory complicates analysis. This is the main reason sufficient assumptions leading to optimal contracts without memory are sought after. If we consider (as Radner does) repeated games in which the contracts have no memory, the repeated game can be played myopically, that is the optimal one-period behavior is also an optimal multiperiod behavior. Sufficient assumptions about such contracts without memory being optimal have been derived by Fellingham, Newman and Suh (1985). The assumptions are rather restrictive in nature.

Summing up the discussion up to that point, it may be noted that we have to be very careful of drawing strong conclusions from the one-period models because of the likely instability of their results within a multiperiod setting. But so far we cannot gain further insight from multiperiod models which allow handling of an (at least) three-person agency relationship.

Criticism of the discussed agency theory models may be expressed further due to the negligence of more than one manager and auditor, respectively. That criticism corresponds with the advice not to disregard markets. It should be remembered that one of the important results of the model of Antle was that coalition-forming of auditor and manager may be expected. This problem may be due to the fact that markets are omitted. It has also been addressed explicitly by Baiman who stated that an allowance for collusive behavior among a more than two-person agency relationship "may give rise to substantially different employment contracts and managerial accounting procedures" than in a two-person relationship (1982, p. 177).

The discussion of that point is similar to the one of the planning horizon in agency models. At least, parts of the problems have been foreseen, and there are some approaches with multiple agents (cf. especially Demski and Sappington 1984, Mookherjee 1984 and Holmström 1982). But those approaches have neither adressed the problem of auditing nor that of coalition-forming. Therefore, the models are not comparable with the models which we have discussed.

Disregarding all formal aspects, the arguments of Watts and Zimmerman could be used to criticize the view that collusive behavior of auditor and manager may be expected. They especially point out that it should be in the best interests of the manager to be audited by an independent auditor, because in any other case the manager has to expect that he has to bear the costs of his opportunistic behavior (cf. Watts and Zimmerman 1983, pp. 614 - 615). Besides this, the auditors should be motivated to signal that they are independent and do their job well. The main reason for such behavior is the auditor's reputation. "If found to have been less independent than expected, the auditor's reputation is damaged and the present value of the auditor's services is reduced. He bears costs." (Watts and Zimmerman 1986, p. 316) In order to reduce the costs of providing owners with information about an auditor's independence (and competence) there could be expected a professional auditors' society that accredits auditors. Such societies which help to gain trust in contracting partners really exist, but they are neglected in the agency models.

6.2 Practical Relevance

Thus far, the models cannot be used to make decisions. If the owner of a firm has to decide on the financial reporting alternatives of the manager, on mobilizing an auditor, on the intensity of auditing (for example, by means of time and money that should be spended and desired quality of people who should be engaged), on the form of the auditor's report or the like, then the models are not feasible to make those decisions easier. They are not constructive in this sense. Admitting this, it should be remembered that this is not unique in the field of economic theory.

Even if there is no decision-supporting function of the models, they are not superfluous. They tell us that even in a world with very favorable assumptions about the knowledge of a principal with respect to the properties of the decision problems and the personal attributes of his agents, he cannot neglect coalition-forming if he tries to motivate one agent in order to motivate another one. The conditions which are sufficient for a positive value of auditing must therefore be expected to be rather restrictive. This conclusion, of course, may be gained with devices other than agency theory, too. The exact conditions still await further research since we have seen, that the results of one-period three-person agency models may change if the assumptions are generalized to more than one period and more than three persons.

References:

Antle, R.L. (1981): Moral Hazard and Auditor Contracts: An Approach to Auditors' Legal Liability and Independence, Ph.D. Diss., Stanford University

Antle, R.L. (1982): The Auditor as an Economic Agent, Journal of Accounting Research 20, 503-527

Arrow, K.J. (1985): The Economics of Agency, in: Pratt, J.W.; Zeckhauser, R.J. (eds.): Principals and Agents: The Structure of Business, Boston, Mass., 37-51 (almost identical with Agency and the Market, in: Arrow, K.J.; Intriligator, M.D. (eds.): Handbook of Mathematical Economics, Vol. III, Amsterdam 1986, 1183-1195)

ASOBAC (1972): Committee on Basic Auditing Concepts, A Statement of Basic Auditing Concepts, Accounting Review 47, Supplement, 15-74

Bailey Jr., A.D. (1981): Statistical Auditing: Review, Concepts, and Problems, New York, London, Sydney etc.

Baiman, S. (1979): Discussion of Auditing: Incentives and Truthful Reporting, Journal of Accounting Research 17, Supplement, 25-29

Baiman, S. (1982): Agency Research in Managerial Accounting: A Survey, Journal of Accounting Literature 1, 154-213 (reprinted in Mattessich, R. (ed.): Modern Accounting Research: History, Survey, and Guide, Vancouver 1984, 251-294)

Baiman, S.; Evans III, J.H.; Noel, J. (1985): Optimal Contracts with a Utility-Maximizing Auditor, Unpublished Working Paper, Graduate School of Industrial Administration, Carnegie-Mellon University, and Graduate School of Business, University of Pittsburgh

Ballwieser, W. (1985): Ergebnisse der Informationsökonomie zur Informationsfunktion der Rechnungslegung, in: Stöppler, S. (ed.), Information und Produktion, Essays in Honor of W. Wittmann, Stuttgart, 21-40

Ballwieser, W. (1987): Kapitalmarkt, Managerinteressen und Rolle des Wirtschaftsprüfers, in: Schneider, D. (ed.): Kapitalmarkt und Finanzierung, Berlin, in print

Demski, J.S.; Patell, J.M.; Wolfson, M.A. (1984): Decentralized Choice of Monitoring Systems, Accounting Review 59, 16-34

Demski, J.S.; Sappington, D. (1984): Optimal Incentive Contracts with Multiple Agents, Journal of Economic Theory 33, 152-171

Fama, E.F. (1980): Agency Problems and the Theory of the Firm, Journal of Political Economy 88, 288-307

Fellingham, J.C.; Newman, D.P.; Suh, Y.S (1985): Contracts without Memory in Multiperiod Agency Models, Journal of Economic Theory 37, 340-355

Harris, M.; Raviv, A. (1978): Some Results on Incentive Contracts with Applications to Education and Employment, Health Insurance, and Law Enforcement, American Economic Review 68, 20-30

Harris, M.; Raviv, R. (1979): Optimal Incentive Contracts with Imperfect Information, Journal of Economic Theory 20, 231-259

Holmström, B. (1979): Moral Hazard and Observability, Bell Journal of Economics 10, 74-91

Holmström, B. (1982): Moral Hazard in Teams, Bell Journal of Economics 13, 324-340

Lambert, R.A. (1983): Long-Term Contracts and Moral Hazard, Bell Journal of Economics 14, 441-452

Levinthal, D.A. (1985): Three Essays on Economic Models of Organizations, Ph.D. Diss., Stanford University

Mookherjee, D. (1984): Optimal Incentive Schemes with Many Agents, Review of Economic Studies 51, 433-446 (reprinted in Binmore, K.; Dasgupta, P. (eds.): Economic Organizations as Games, Oxford, New York 1986, 197-214)

Myerson, R.B. (1979): Incentive Compatibility and the Bargaining Problem, Econometrica, 61-73

Namazi, M. (1985): Theoretical Developments of Principal-Agent Employment Contracts in Accounting: The State of the Art, Journal of Accounting Literature 4, 113-163

Ng, D.S. (1978): An Information Economics Analysis of Financial Reporting and External Auditing, Accounting Research 53, 910-920

Ng, D.S.; Stoeckenius, J. (1979): Auditing: Incentives and Truthful Reporting, Journal of Accounting Research 17, Supplement, 1-24

Noel, J.C. (1981): Agency Costs and the Demand and Supply of Auditing, Ph.D. Diss., The Ohio State University

Penno, M. (1985): Informational Issues in the Financial Reporting Process, Journal of Accounting Research 23, 240-255

Radner, R. (1981): Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship, Econometrica 49, 1127-1148

Rees, R. (1985): The Theory of Principal and Agent. Part 1 and 2, Bulletin of Economic Research 37, 3-26 and 75-96

Roberts, D.M. (1978): Statistical Auditing, New York

Rogerson, W.P. (1985): Repeated Moral Hazard, Econometrica 53, 69-76

Shavell, S. (1979): Risk Sharing and Incentives in the Principal and Agent Relationship, Bell Journal of Economics 10, 55-73

Townsend, R.M. (1982): Optimal Multiperiod Contracts and the Gain from Enduring Relationships under Private Information, Journal of Political Economy 90, 1166-1186

Verrecchia, R.E. (1986): Managerial Discretion in the Choice among Financial Reporting Alternatives, Journal of Accounting and Economics 8, 175-195

Watts, R.L.; Zimmerman, J.L. (1983): Agency Problems, Auditing, and the Theory of the Firm: Some Evidence, Journal of Law and Economics 26, 613-633

Watts, R.L.; Zimmerman, J.L. (1986): Positive Accounting Theory, Englewood Cliffs, N.J.

Woodland, B.M. (1981): Incentive Contracts with Auditing and Monitoring in the Principal-Agent Relationship, Ph.D. Diss., Purdue University

Yandell, D.S. (1981): Auditing in a Principal-Agent Model: Information and Incentives in Contract Design, Ph.D. Diss., Purdue University

Investigation Strategies with Costly Perfect Information

Alfred Wagenhofer

Summary: An agency model where the principal gets no information at all, but can acquire perfect information at a constant cost, is considered. The principal has to decide which investigation strategy he can optimally precommit to follow. In the case of no communication, low cost, and insufficient penalty, preferability of a random investigation strategy is shown. In the case of communication of the agent's superior information, probabilities of investigation decreasing in reported outcome are optimal.

Table of Contents:

1. Introduction

Many results on the optimal design of contracts between principal and agent are available for costless information. But considering costs of information may have a strong impact on the validity of those results.

I would like to thank Professors Günter Bamberg, Martin Hellwig, the participants in a workshop at the University of British Columbia, and Dieter Kahr for helpful comments.

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

For instance, the principal may do better without acquiring information. This paper analyzes the effects of the cost of perfect information on the information acquisition or investigation strategy employed by the principal. The model incorporates the extreme cases of the principal's information which are

(i) no information at all, and
(ii) perfect but costly information about both the agent's action and the output produced.

The first objective of this paper is to identify conditions for random investigation to be optimal if no communication is possible. It turns out that if the cost of investigation is "low" (in a sense which will be made clear later) it is optimal to always investigate, otherwise not investigating at all is the optimal strategy. Random investigation occurs if the principal does not have sufficient penalty available to enforce the agent to choose the optimal action. The probability of investigation then depends on that penalty available and *not* on the cost of investigation.

The second objective is to see how communication changes this result. Before the principal decides if to investigate he asks the agent to tell him which output had occurred, and which action he had chosen. The basic result is that communication is weakly preferable (provided it is costless), but strict gains arise in general. Strict gains occur by utilizing the availability of sufficient penalty which can be imposed upon the agent. Though the limiting result is trivial (viz. penalty without lower bounds and probability of investigation going to zero approach the first-best contract arbitrarily closely) for exogenously specified penalty it is shown that the optimal investigation strategy is decreasing in reported output. The decreasing probabilities come from both guaranteeing selection of the agreed upon action, and truthful revelation of the outcome by the agent. If only insufficient penalty is available, a less strong result shows that there exists a strictly decreasing probability function which provides a lower bound on the optimal probability of investigation.

A paper investigating a related issue is by Townsend [1979], who analyzes risk-sharing contracts with costly conditional information. Evans [1980] extends Townsend's analysis by moral hazard on the part of the agent. Both analyze interval investigation contracts, which are contracts splitting the support of possible outcomes into a set of outcomes for which there is always investigation, and the complementary set for which there is no investigation at all. They find that a lower interval investigation is strictly preferable to any other interval investigation, investigation being carried out with the same probability mass. However, Townsend presents an example which shows strict preferability of decreasing probabilities in output.

There are some other papers dealing with this type of issue. Demski and Feltham [1978] show a strict Pareto-improvement by allowing for costly lower interval investigation of the action taken conditional on the output, which they assume to be public information. They provide an example with different levels of penalty, but consider only sufficient penalty.

In a recent study Mookherjee and Png [1986] show the optimality of decreasing probabilities of investigation conditional on reported outcome in a taxation setting which can be interpreted in terms of the model underlying this paper. However, their results regard only cases in which the ill-informed government cannot observe the action. This paper differs from theirs mainly in this assumption which is shown to have an impact on the results. In another context Kanodia [1985], too, derives decreasing probabilities by considering moral hazard by the agent. His result is especially due to the assumption of the agent having predecision information about the environment.

Baiman and Demski [1980a, 1980b], and subsequently Lambert [1985] analyze situations of performance evaluation in which the principal always observes the output and can costly investigate to get imperfect information about the action or the environment. Their results show that always investigating or not investigating at all for a certain outcome is optimal. They also show that while a lower interval investigation is optimal for "very" risk-averse agents the investigation interval shifts to an upper outcome subset for "not very" risk-averse agents. This result strongly depends on the assumption of the utility functions for, as Young [1986] shows, a two-tailed investigation region arises in the same model if one allows for another utility class than that considered by Baiman and Demski. Recently Dye

[1986] extended the model to include perfect information about the action, showing the optimality of lower interval investigation for any strictly concave utility function.

The paper proceeds as follows. In section 2 the model with costly conditional information is introduced. The main result concerning the preferability of a random investigation strategy is derived in section 3, where an example is provided as well. Communication is considered in section 4, and the results for sufficient penalty available are derived in section 5. Section 6 contains a characterization of the investigation strategy for insufficient penalty, and illustrates the results by continuing with the example. The conclusions appear in section 7.

2. The Model without Communication

A principal hires an agent to perform some action unobservable to the principal which influences the output of a given productive process.

Let $a \in \mathbf{A} = [\underline{a}, \bar{a}]$ denote the action chosen by the agent causing effort. $\Theta \in \mathbf{\Phi}$ be an exogenous random state variable. Output in terms of money is denoted by $x(a,\Theta) \in [\underline{x}, \bar{x}]$. Let $s(x,a)$ be the share of the output to be transferred to the agent for supplying the action.

Both the principal (P) and the agent (A) behave as if they were expected utility maximizers, their utilities can be described by von Neumann-Morgenstern utility functions with the usual properties. Let $G(x)$ be the utility function of P and assume P to be risk-neutral, i.e. $G(x) = x$. Be $H(s,a)$ the utility function of A and additively separable in s and a , $H(s,a) = U(s) - V(a)$. Let A be strictly risk-averse in s and weakly effort averse ($H_s(s,a) > 0$; $H_{ss}(s,a) < 0$; $H_a(s,a) < 0$; $H_{aa}(s,a) \leq 0$).[1]

[1] Subscripts denote the partial derivative with respect to the argument listed. If there is only one argument in a function, the derivative is denoted by a prime.

P and A have homogeneous expectations about the random state variable Θ prior to agreeing to the contract. The riskiness of Θ will be incorporated in a parameterized conditional distribution function $F(x|a)$ with the density function $f(x|a) > 0$ (for all x), $f_a(x|a)$ and $f_{aa}(x|a)$ may exist. Assume moreover that the MLRP (monotone likelihood ratio property, that is $f_a(x|a)/f(x|a)$ is increasing in x) and CDFC (convexity of the distribution function condition, that is $F_{aa}(x|a) \geq 0$ all x, a) hold for $F(x|a)$ to assure validity of the first-order approach (see Rogerson [1985]).[2] Assume existence of the solutions to the model described and existence of an interior optimal action.

P can investigate at constant cost $C > 0$, the investigation once conducted provides perfect information about the x and a having occurred.[3] Since P has to decide if to investigate before acquiring any knowledge of x or a the probability $\alpha \in [0,1]$ of conducting an investigation must be a *constant* (i.e., independent of x or a). A wheel is spun, after the agent chose his action, but before the compensation is paid. The result is either "investigate" or "do not investigate".

If an investigation takes place P gets all the information that A has. Then the optimal compensation scheme is the *pure wage contract* (PWC) described by the set of variables $\{s_w(a), a_w\}$ which gives the so called first-best or cooperative solution.

$$s_w(x,a) = s_w(a) = \begin{cases} s_w & \text{if } a = a_w \\ \underline{s} & \text{else .} \end{cases} \tag{1}$$

[2] MLRP implies that the action a shifts $F(x|a)$ to the right in the sense of first-order stochastic dominance (i.e. $F_a(x|a) \leq 0$ for all x). CDFC implies stochastically decreasing returns to scale in a i.e. $x_a(a,\Theta) > 0$ and $x_{aa}(a,\Theta) < 0$.

[3] Baiman [1979] pointed out the difference between monitoring the action, and auditing the outcome. Whereas auditing can be conducted after the production process has ended, monitoring can be done only during the action being performed. To allow for a decision about monitoring afterwards it could be assumed that monitoring always occurs but the decision is about costly evaluation ex post.

That means, since P is risk-neutral he absorbs all the risk associated with x and pays A a constant share as long as A chooses the agreed upon action. Otherwise a penalty $\underline{s}$ is imposed on A.

On the other hand, if no investigation is performed the only feasible contract is the *pure rental contract* (PRC) characterized by $\{k_r, a_r\}$. The reason is that P is not able to observe the actual x or a having occurred hence any contract contingent on these variables is subject to moral hazard by the agent. The optimal compensation scheme therefore is

$$s_r(x,a) = s_r(x) = x - k_r \quad . \tag{2}$$

P gets a constant k_r and A bears all the risk of the output which implies that no risk-sharing takes place.

The program to derive the optimal contract with $\{s(a), k, a, \alpha\}$ therefore is of the following form:

$$\max_{\alpha,s,k} EG = \alpha\int xf(x|a)dx - \alpha s(a) - \alpha C + (1-\alpha)k \tag{3}$$

subject to

$$EH = \alpha U(s(a)) + (1-\alpha)\int U(x-k)f(x|a)dx - V(a) \geq H_m \tag{4}$$

$$a \; \varepsilon \; \operatorname{argmax}_{a'} \{\alpha U(s(a'))+(1-\alpha)\int U(x-k)f(x|a')dx - V(a')\} \quad \text{(ASC)} \tag{5}$$

Equation (3) is the maximization problem for P, and (4) is the *market constraint* for A to agree to the contract. (5) is the *action selection constraint* (ASC) which states that the action A chooses is one of the actions most preferable to himself.

The time sequence is summarized in Figure 1.

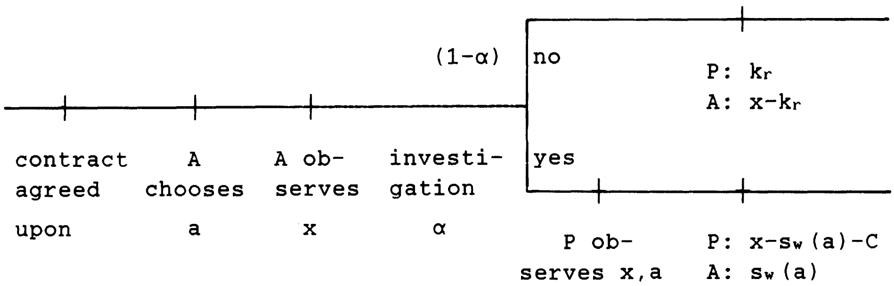


Figure 1: Time line for the basic model

Definition: A penalty $\underline{s}'$ is said to be *sufficient* (sufficiently low) to enforce the PWC if $\underline{s}' \leq \underline{s}$, where $\underline{s}$ is defined by

$$H_m = U(s_w) - V(a_w) = \max_a \{U(\underline{s}) - V(a)\} = U(\underline{s}) - \min_a V(a)$$

or equivalently,

$$U(\underline{s}) = H_m + V(\underline{a}) \ . \tag{6}$$

Proposition 1:

(i) If

$$C < \int xf(x|\, a_w)dx - s_w - k_r \tag{7}$$

and if $\underline{s}'$ is sufficient then $\alpha = 1$.

(ii) If $C > \int xf(x|\, a_w)dx - s_w - k_r$ then $\alpha = 0$.

(iii) If the condition holds as an equality then $\alpha = 0$ is optimal; and if $\underline{s}'$ is sufficient then $\alpha = 1$ is also optimal.

Proof:
The Lagrangian function corresponding to the maximization program, holding the action constant, is linear in α which implies that the optimal α is a corner solution; that is $\alpha \in \{0,1\}$. Either $\alpha = 0$ or $\alpha = 1$ must be optimal. Since this is valid for any action, it must hold for the optimal action as well (for the same line of argument see Baiman and Demski [1980a, Proposition 3]).

It then suffices to compare the expected utilities generated by these two extreme cases.

If $\alpha = 1$ then $EH = H_m$ and $EG = \int xf(x|a_w)dx - s_w - C$.

If $\alpha = 0$ then $EH = H_m$ and $EG = k_r$. Observe that in both cases A can be held to his reservation utility.[4] Now, of course, $\alpha = 1$ is optimal if

$$k_r < \int xf(x|a_w)dx - s_w - C$$

or equivalently,

$$C < \int xf(x|a_w)dx - s_w - k_r ,$$

if there is sufficient penalty available. This was stated in the Proposition.

If C is greater than the right hand side of (7) $\alpha = 0$ will be the optimal investigation policy. C equal to the right hand side of (7) yields the same expected utilities for $\alpha \in \{0,1\}$ given $\underline{s}'$ is sufficient. Any other $\alpha \in (0,1)$ cannot be optimal since A chooses a different action for $\alpha = 0$ and $\alpha = 1$, respectively. If P would precommit to an $\alpha \in (0,1)$, A could not avoid being penalized with some probability thus lowering his expected utility.

Q.E.D.

That is, if the cost of investigation is sufficiently high it is optimal to *never* investigate. Otherwise if the cost is sufficiently low an investigation policy with *always* investigating is optimal provided that there exists sufficient penalty. Then randomization is not preferable to pure investigation strategies. This result is similar to that derived in Baiman and Demski [1980a], and follows from the fact, that communication is excluded. The additional information (be it an imperfect monitor of the action, as in Baiman and Demski, or the outcome and action, as in this paper) cannot be distorted by the agent. Later in this paper we will see that otherwise the optimal investigation strategy generically is not bang-bang.

[4] If this were not true, a new contract could be constructed by lowering the compensation of A for all outcomes without changing the incentives, thus increasing EG .

3. Occurrence of Random Investigation

The key for the result above is the availability of sufficient penalty. The question now is what happens if the penalty is not sufficient. Then the first-best solution is not attainable since A would be better off choosing his lowest action $\underline{a}$ making himself better off ($EH > H_m$) and P worse off. Denote the available penalty by $\underline{s}_h$, $\underline{s}_h > \underline{s}$. The only enforceable contract would be the PRC which always works since it needs no enforcement. But the following result shows that P can do better under certain conditions by employing a mixed strategy, by offering a *random investigation contract* (RIC) described by $\{s_0(a), k_0, a_0, \alpha_0\}$, where $\alpha_0 \in (0,1)$.

Proposition 2: Assume C so small that (7) holds, and $\underline{s}_h > \underline{s}$. Then an RIC is strictly preferable to the PRC if the following is true:

$$\underline{s}_h < s_r \ , \qquad (8)$$

where $U(s_r) = \int U(x-k_r) f(x| a_r) dx$,

and one of the following three conditions holds:

(i) $R \equiv EG(a_r, \alpha{=}1) - k_r \geq 0$,

(ii) $R < 0$, *and* $EG_\alpha(a_r, \alpha_r{=}0) \geq 0$,

(iii) $R < 0$, $EG_\alpha(a_r, \alpha_r{=}0) < 0$, *and* $\alpha_0 > \alpha'$,
where $0 < \alpha' < 1$ *is defined by* $EG(a', \alpha') = EG(a_r, \alpha_r{=}0)$.

Proof:

Given an α_0 the optimal $s_0(a)$, k_0 , a_0 can be derived from the program (3) to (4) without the ASC (5). This contract conditional on α_0 can be compared to the PRC, and the conditions for the preferability of the RIC follow. In a second step the necessary penalty $\underline{s}_h$ to enforce that contract (i.e. to satisfy the ASC given $\{s_0(a), k_0, a_0$ $\alpha_0\}$) is derived. This gives the necessary condition (8). It suffices to compare the $EG(\cdot)$ since $EH = H_m$ for all the contracts considered.

By the assumption of C so small that (7) holds, we know from Proposition 1 that

$$EG(a_w, \alpha_w{=}1) - k_r > 0 \ ,$$

where $k_r = EG(a_r, \alpha_r=0)$, i.e., P prefers the PWC but cannot enforce it. Notice that the contracts $\{s_w(a), a_w, \alpha_w=1\}$ and $\{k_r, a_r, \alpha_r=0\}$ are the optimal contracts for the extreme values of α. The gains from the PWC consist of improved risk-sharing since if an investigation is conducted, risk is shifted from A to P by the constant compensation in this case. Gains, too, result from selecting a higher action in the case of investigating.

Lemma 1: $a_r < a_0 < a_w$ *for* $0 < \alpha_0 < 1$, *and* $\frac{da_0}{d\alpha_0} > 0$.

Proof:
First it is shown that $a_r < a_w$. The first-best action a_w is determined by

$$\max_a EG = \int x f(x|a)dx - s_w(a) \tag{3'}$$

s.t.

$$U(s_w(a)) - V(a) \geq H_m \tag{4'}$$

The optimal action a_r for the PRC comes from solving

$$\max_k EG = k \tag{3''}$$

s.t.

$$\int U(x-k)f(x|a)dx - V(a) \geq H_m \tag{4''}$$

$$\int U(x-k)f_a(x|a)dx - V'(a) = 0 . \tag{5''}$$

Suppose we extract all risk-sharing gains in (3') and (4') by setting $x-k_r$ for s_w we get the equations (3") and (4"). Now (5") is always binding, as it is the ASC for the problem for A. So (5") must restrict the possible solutions for a_r in the PRC from above (we have assumed an interior solution). We therefore conclude $a_w > a_r$.[5] Now $a_0 = a_r$ for $\alpha_0 = 0$, and $a_0 = a_w$ for $\alpha_0 = 1$.

Next we show that for $a_0 > a_r$ $\frac{da_0}{d\alpha_0} > 0$.

[5] The only case that $a_w = a_r$ arises if A is risk-neutral. Then the PRC gives the same utilities as the PWC in case of symmetric information. This can easily be seen by observing that the ASC in (5") then equals the action selection derived from the Lagrangian of problem (3') and (4').

Differentiate the Lagrangian corresponding to (3) and (4) with respect to a, which defines a_0 :

$$T \equiv \frac{dL}{da} = \alpha_0 \int x f_a(x|a_0)dx + \tau(1-\alpha_0)\int U(x-k) f_a(x|a_0)dx - \tau V'(a_0) = 0 .$$

To get the desired result we differentiate T to get:

$\frac{dT}{da_0} < 0$ since this is the second-order condition for a_0 .

$$\frac{dT}{d\alpha_0} = \int x f_a(x|a_0)dx - \tau\int U(x-k) f_a(x|a_0)dx .$$

Replacing by T this becomes

$$\frac{dT}{d\alpha_0} = \frac{\tau}{\alpha_0}\cdot[V'(a_0) - \int U(x-k) f_a(x|a_0)dx] > 0 .$$

The last inequality follows from $\tau > 0$ since the constraint (4) is binding. And the term in brackets is the negative first-order condition for a_r, which is negative for $a_0 > a_r$ since the second-order condition for a_r is negative.

Now $\frac{da_0}{d\alpha_0} = -\frac{dT}{d\alpha_0} / \frac{dT}{da_0}$. Therefore $\frac{da_0}{d\alpha_0} > 0$.

Q.E.D.

Continuing with the proof of Proposition 2, we now prove the conditions for the preferability of the RIC. Consider first the gains from risk-sharing alone as compared to the cost of investigation. They can be calculated by fixing the action a_r, and reducing risk imposed on A through the compensation scheme. The maximal net gains from risk-sharing are

$$R \equiv EG(a_r, \alpha=1) - k_r . \quad (9)$$

Suppose $R > 0$, then varying α within $[0,1]$ creates a linear combination resulting in a strict improvement of $EG(\cdot)$ for any $\alpha > 0$, since a_r can be enforced at any case since $a_0 > a_r$. But the last argument shows that also in the case $R = 0$ strict gains will occur since EG increases with a higher action. This establishes condition (i) of the Proposition.

If $R < 0$ then the functional form of $EG(\cdot)$ must be regarded as well. A variation of α_0 generates a smooth function $EG(a_0, \alpha_0)$ starting from the point $EG(a_r, \alpha_r = 0)$ to $EG(a_w, \alpha_w = 1)$. $EG_\alpha(a_r, \alpha_r = 0) \geq 0$ and the fact that $EG(\cdot)$ increases in α imply $EG(a_0, \alpha_0) > EG(a_r, \alpha_r = 0)$ for all $\alpha_0 > 0$, which proves condition (ii). Observe that condition (i) is sufficient for $EG_\alpha(a_r, \alpha_r = 0) > 0$.

If $EG_\alpha(a_r, \alpha_r = 0) < 0$ then there exist some α_0 for which $EG(a_0, \alpha_0) \leq EG(a_r, \alpha_r = 0)$, which means that the PRC is preferable. But since $EG(\cdot)$ increases in α , and $EG(a_w, \alpha_w = 1) > EG(a_r, \alpha_r = 0)$ by the assumption for C , there exists one α' , $0 < \alpha' < 1$, such that $EG(a', \alpha') = EG(a_r, \alpha_r = 0)$. For all $\alpha_0 \in (0, \alpha')$ $EG(a_0, \alpha_0) < EG(a_r, \alpha_r = 0)$, and for all $\alpha_0 \in (\alpha', 1)$ $EG(a_0, \alpha_0) > EG(a_r, \alpha_r = 0)$. This proves condition (iii).

To show the need for condition (8), $\underline{s}_h$ must be sufficiently low to assure that A will choose the agreed action a_0 . Hence we construct $\underline{s}_h$ such that the ASC is satisfied which is

$$\alpha_0 U(s_0) + (1-\alpha_0)\int U(x-k_0) f(x|a_0)dx - V(a_0) \geq$$

$$\max_{a'} \{\alpha_0 U(\underline{s}_h) + (1-\alpha_0)\int U(x-k_0) f(x|a')dx - V(a')\} .$$

Since A can be held to his reservation utility H_m , the constraint can be rewritten as

$$\alpha_0 U(\underline{s}_h) \leq H_m - (1-\alpha_0)\int U(x-k_0) f(x|a_1)dx + V(a_1) ,$$

where a_1 maximizes $(1-\alpha_0)\int U(x-k_0) f(x|a)dx - V(a)$.

The (highest) necessary penalty to enforce contract $\{s_0, k_0, a_0, \alpha_0\}$ then is

$$U(\underline{s}_h) = \frac{1}{\alpha_0} \{H_m - (1-\alpha_0)\int U(x-k_0) f(x|a_1)dx + V(a_1)\} . \qquad (10)$$

It is very difficult to explicitly show that α_0 increases if more penalty is available. The reason is that (10) contains an implicit maximization problem. But we can find upper and lower values for $\underline{s}_h$ that induce an RIC.

$$U(\underline{s}_h)\Big|_{\alpha=1} = \lim_{\alpha\to 1} \frac{1}{\alpha} \{H_m - (1-\alpha)\int U(x-k_0)f(x|a_1)dx + V(a_1)\}$$

$$= H_m + V(a_1)\Big|_{\alpha=1} = H_m + V(\underline{a}) \ .$$

By (6) it follows that $\underline{s}_h\Big|_{\alpha=1} = \underline{s}$.

$$U(\underline{s}_h)\Big|_{\alpha=0} = \lim_{\alpha\to 0} \frac{1}{\alpha} \{H_m - (1-\alpha)\int U(x-k_0)f(x|a_1)dx + V(a_1)\}$$

$$= \lim_{\alpha\to 0} \frac{1}{\alpha} \{\int U(x-k_r)f(x|a_r)dx - V(a_r) - (1-\alpha)\int U(x-k_r)f(x|a_r)dx + V(a_r)\}$$

$$= \lim_{\alpha\to 0} \frac{1}{\alpha}\, \alpha\int U(x-k_r)f(x|a_r)dx = \int U(x-k_r)f(x|a_r)dx = U(s_r) \ .$$

An RIC can only prevail if $\underline{s} < \underline{s}_h < s_r$, which is assured by the initial assumption to Proposition 2 and equation (8).

Q.E.D.

The necessary condition (8) says that the penalty inducing random investigation must be lower than the certainty equivalent of the agent's compensation in the case of the PRC. Otherwise the only enforceable contract is the PRC.

The intuition behind the other conditions is that as long as the cost of investigation is lower than the benefits due to enforcing the PWC there exists a region within the admissible $\underline{s}_h$ (and corresponding α_0) where a combination of this contract with the PRC must be preferred to the PRC alone. The PWC would provide more than A's reservation utility, therefore the PRC is constructed to provide less, and randomizing gives exactly the reservation utility. Although random investigation seems to impose more risk upon A it actually takes away risk by allowing for the first-best compensation scheme in the case of investigation. The necessary penalty is constructed so that the RIC under consideration can be enforced. The benefits of the RIC arise from improved risk-sharing and from agreeing to a higher action than under the PRC but lower than the first-best action (which would not prevail because of insufficient penalty). The preferred region extends over the full support of α in the case that the cost is lower than the gains to improved risk-sharing alone. But as long as the increase of P's expected utility at $\alpha = 0$ is greater or equal zero, which is also assured if this curve is sufficiently concave, then there exists no region where the PRC is optimal. Otherwise, especially in cases in which $EG(a_0,\alpha)$ is convex in α, there is a region of small α where the PRC remains optimal though P would prefer to investigate.

It may be interesting to say something about the condition of $EG(a_0,\cdot)$ being a concave or a convex function.

$$\frac{dEG}{d\alpha} = \int xf(x|a_0)dx + \alpha\int xf_a(x|a_0)\frac{da_0}{d\alpha}dx - s - k - C \quad .$$

$EG(a_0,\alpha)$ is convex if $\frac{d^2EG}{d\alpha^2} > 0$. Now

$$\frac{d^2EG}{d\alpha^2} = 2\int xf_a(x|a_0)\frac{da_0}{d\alpha}dx + \alpha\int xf_{aa}(x|a_0)\frac{da_0}{d\alpha}dx + $$
$$+ \alpha\int xf_a(x|a_0)\frac{d^2a_0}{d\alpha^2}dx \quad .$$

The first two terms are positive (by first-order stochastic dominance), and the sign of $\frac{d^2a_0}{d\alpha^2}$ determines the sign of the third term.

Presumably $EG(a_0,\cdot)$ therefore generically is a convex function in α which means that in the case of higher cost than benefits from risk-sharing there is a lower probability interval, and an upper penalty interval for which random investigation is *not* optimal.

An interesting fact is that the optimal investigation probability does not directly depend on C which would be an intuitive conjecture, but is fixed by the available penalty $\underline{s}_h$ instead.

An *example* will be provided for illustrating the results of Proposition 2. Consider the following situation:

$$0 \le x \le \infty \ , \ 0 < a < \infty \ , \quad f(x|a) = \frac{1}{a}e^{-x/a}$$

$$U(s) = -e^{-s} \ , \quad V(a) = a, \quad H_m = -2.2974 \ .$$

Although the assumptions do not fit exactly (the support of x is not finite) results can be obtained with quite simple calculation.

a	α	EG(C=0)	EG(C=0.1)	EH	$\underline{s}$, $\underline{s}_h$	s	k
a_r = 0.6487	0	1.0	1.0	-2.2974	-	-	1.0
a_0 = 0.8315	0.3	1.0555	1.0255	-2.2974	-0.5851	-0.3824	0.9876
a_0 = 0.9707	0.5	1.1072	1.0572	-2.2974	-0.6583	-0.2827	0.9611
a_0 = 1.1100	0.7	1.1727	1.1027	-2.2974	-0.7500	-0.1718	0.9185
a_w = 1.2974	1	1.2974	1.1974	-2.2974	-0.8318	0.0	-
a_r = 0.6487	1	1.1487	1.0487	-2.2974	-0.8318	-0.5	-

Table 1: Some Results of the Example

The maximal difference due to risk-sharing and inducing the optimal action is $EG(a_w, \alpha_w=1, C=0) - EG(a_r, \alpha_r=0) = 0.2974$ (comparing gross expected utilities). This determines the maximum cost C for always investigating to make P better off. The gains to pure risk-sharing are $EG(a_r, \alpha=1, C=0) - EG(a_r, \alpha=0) = 0.1487$ which is the limit for the size of C such that random investigation always makes P better off than the PRC. In this case $EG(\cdot)$ is convex in α such that for $0.1487 < C < 0.2974$ there exists a lower interval where the pure rental contract remains optimal. The different cases are depicted in Figure 2. As can be seen, for the case $C = 0.2$ a probability of investigation α_0 from 0 to about 0.404 leads to less expected utility for P than not conducting an investigation.

The availability of penalty is the crucial parameter in the model. It determines if random investigation occurs, and how the investigation probability is set. Looking at the penalty more closely there turn out to arise some issues regarding the interpretation of this penalty.

There are two different interpretations of the RIC as a combination of the PWC and the PRC. One is that in case of the PRC A is in the position of an entrepreneur and in case of the PWC A serves as employee. The second is thinking of A always acting as entrepreneur, although he may be paid a fixed wage. A penalty is necessary in the PWC to ensure that A will do what P wants him to do. Observe that the penalty will *never* be imposed. It is only a threat since it is A's decision to choose that action precluding penalty.

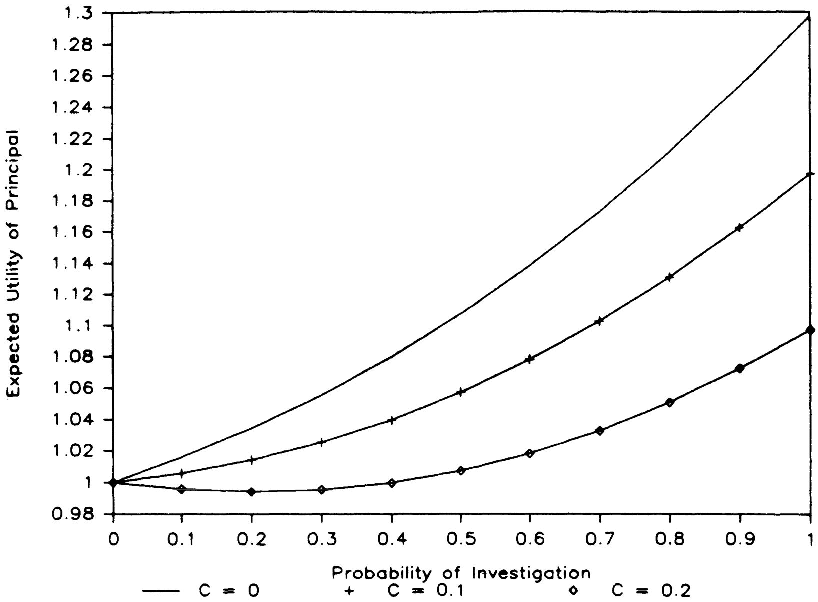


Figure 2: Expected utility of principal at various costs of investigation

In reality it often happens that the penalty which can be imposed on employees is restricted but there are no restrictions to losses due to entrepreneur activities (this would state a preference on the employer-employee interpretation). The reason is that the entrepreneur makes his own decisions, is not being monitored, and has the chance of gaining very much as well. Of course we implicitly assume in this case that he is endowed with enough initial wealth to meet any losses.

Now it may happen that the sufficient penalty is higher than the worst results in the rental case such that

$$\underline{s} > \underline{x} - k_r \quad . \tag{11}$$

This could be interpreted as allowing for the imposition of more "penalty" upon A in the case he runs the enterprise himself. And this "penalty" could occur with positive probability for A cannot avoid bad outcomes even if he chooses the highest available action (by the no shifting support assumption). On the other hand, acting under the PWC he can avoid being penalized with certainty. Therefore, if only insufficient penalty is available, (11) holds, and the effective share to A is important (without regarding to the basis from which it draws) then neither any RIC nor the PRC are feasible.[6]

Next consider the point of time when the randomization is carried out. Randomizing between the PRC and the PWC to induce A to take the proper action only works the desired way if it occurs *after* A has chosen his action. Otherwise the solution would collapse since it would not give A the necessary incentives. But randomizing after A has carried out the action introduces some problems in the employer-employee case regarding the legal institutions. Which contract is agreed upon at the time A chooses his action? This cannot be answered sufficiently.

The model does not give results for a comparison to the second-best contract (e.g., Holmström [1979]). Observe that the lowest share for A under the second-best is higher than under the PRC. The main result till now in Proposition 2 just states the conditions for a random investigation contract to be strictly preferable to a deterministic investigation contract. A comparison of this result with the second-best contract is not easy since it has to compare results of different maximizations. But I conjecture that the second-best contract in general is not preferable to the RIC. An intuitive explanation for this would be that the second-best contract always fully investigates incurring the cost C but does not make any use of observing the action perfectly. So it foregoes information which is used by the RIC.

4. The Value of Communication

Now since the conditions for random investigation have been derived we pursue the question if allowing for communication leads to a Pareto-superior solution than that of the model considered above. Communication consists of P asking A, who already has observed the actual x, to costlessly tell P which x had occurred and which a he had chosen. The communicated values of x and a are denoted by x_m and

[6] Feasibility is used in the sense that A could be held down to his reservation utility. For an analysis of contracts with ex post constraints on the share of the agent see Sappington [1983].

a_m . In the model without communication P chooses a constant α since no information is available to him at the time of that decision. Knowing x_m and a_m gives him the opportunity to make α contingent on that information such that $\alpha = \alpha(x_m, a_m)$.

Lemma 2: In the optimal contract with communication $\alpha = \alpha(x_m)$ *, i.e.* α *does not depend on the action* a_m *reported by the agent.*

Proof:
P precommits to penalize A if he observes that A has chosen another than the agreed upon action a . Asked to tell which action A has chosen, optimal behavior of A consists of always telling that $a_m = a$. For P could penalize him without carrying out an investigation if that was not the case. This means that P has no advantage to have α contingent on a_m since a_m has no value.

Q.E.D.

A contract including x_m must take into account that this information need not be true but A may have the incentive to lie and make himself better off. This can happen since deviating from the actual x will not be detected if no investigation is carried out. In that situation we can apply the *revelation principle* (e.g., Myerson [1979]), since P precommits how to use the information to be revealed by A. Therefore A need not fear that P exploit his truth-telling behavior. By the revelation principle we can restrict our search for optimal contracts to truth-inducing contracts without loss of generality. This restriction can be formulated by introducing an incentive-compatible constraint into the model. It will be referred to as the *message selection constraint* (MSC).

The MSC assures that A will be better off telling the true x . The best A can do otherwise is to report the x_m maximizing his possible gains if not being detected and the penalty imposed upon him if an investigation is carried out. The gains to lying are $s(x_m)+x-x_m$ in the case of no investigation since A would get the compensation $s(x_m)$ and in addition could collect the difference between the actual outcome and the reported outcome.

The time sequence of the events is depicted in Figure 3.

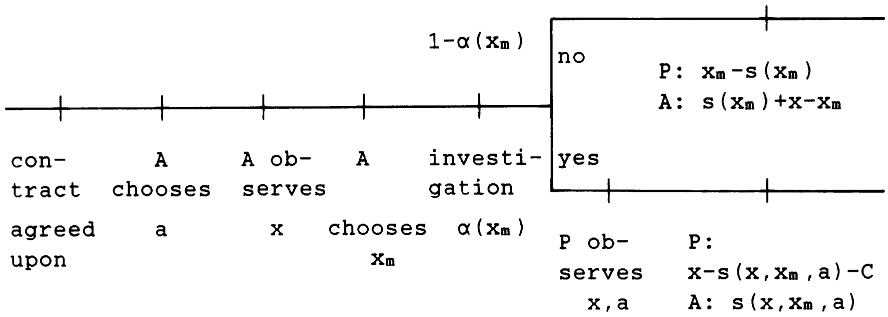


Figure 3: Time line for the communication model

Denote the available penalty by $\underline{s}'$. The compensation scheme $s(x,a)$ can be split such that in cases where the actual a is found to be different from the agreed action, $s(x,a) = \underline{s}'$, otherwise $s(x,a) = s(x)$.

Observe that the MSC allows only for $s(x)$ and $\alpha(x)$ such that truth-telling is the best response by A. By virtue of the MSC the other equations of the program are simplified by only considering the actual x , and not the reported x_m . From imposing the ASC we further can restrict to consider $s(x)$. The program therefore can be stated as follows.

$$\max_{\alpha(x),s(x)} EG = \int [x-s(x)-\alpha(x)C]f(x|a)dx \tag{12}$$

subject to

$$EH = \int U(s(x))f(x|a)dx - V(a) \geq H_m \tag{13}$$

$$\int U(s(x))f(x|a)dx - V(a) \geq \max_{a'} \{U(\underline{s}')\int \alpha(x)f(x|a')dx + \int (1-\alpha(x))U(s(x))f(x|a')dx - V(a')\} \qquad \text{(ASC)} \tag{14}$$

$$U(s(x)) \geq \max_{x_m} \{\alpha(x_m)U(\underline{s}') + (1-\alpha(x_m))U(s(x_m)+x-x_m)\} \quad \text{for all } x . \qquad \text{(MSC)} \tag{15}$$

To start with, we can give a well known result of the preferability of communication.

Proposition 3:
An investigation contract allowing for communication is weakly Pareto-superior to a contract without communication.

Proof:
The proof is obvious since P can always precommit to ignore the information about x provided by A, and construct the same contract as without communication. Of course, this contract is incentive-compatible since it does not give A any incentive to lie for his information will not be used to pay him. Thus the solution of the model without communication is included in the candidates for a solution to the model considered now.

Q.E.D.

Now this result is of little value if one cannot show strict gains to communication. This will be the purpose of the following two sections. We will distinguish between the cases of more penalty available than sufficient (i.e., $\underline{s}_l < \underline{s}$) and the case of insufficient penalty (i.e., $\underline{s}_h > \underline{s}$, which we dealt with in Proposition 2). The case of a penalty of exactly $\underline{s}$ allows only for the PWC and there cannot be strict gains by using any other contract.

The reason for the distinction lies in the fact that it is comparably easy to state results for sufficient penalty since the compensation scheme has a specific form. For insufficient penalty I was not able to explicitly derive the optimal compensation function, but I will show that the solution is somewhat related to that of sufficient penalty. The example introduced earlier in the paper will be continued and used to illustrate the results of communication.

5. Communication and Sufficient Penalty

The structure of the program without possibility of communication allows only for full investigation in the case of low cost and sufficient penalty. If there is much more penalty available than necessary A could be prevented from shirking even if P would choose an $\alpha < 1$, thus lowering the expected cost of investigation. Whereas P then can force the agent to choose the optimal action he cannot enforce the optimal sharing rule as he can under $\alpha = 1$. This is because if P does not investigate he has no information about the x having occurred and hence cannot have his share depend on the unknown x . This can be overcome by allowing for communication.

Formally, let $\underline{s}_l$ denote the available sufficient penalty, $\underline{s}_l < \underline{s}$, $\underline{s}$ defined in (6). Let A tell P an x_m . Since $\underline{s}_l$ is sufficient, the PWC can be enforced which already incorporates optimal risk-sharing by the compensation scheme $s_w(a)$. Any change away from $s_w(a)$ cannot be preferable. This is the reason why the distinction into sufficient and insufficient penalty was made. Reducing $\alpha(x)$ has only an effect on the expected cost of investigation thus increasing the expected utility of P. In the investigation case the shares are $s_w(a)$ (as defined in (1)) and $x-s_w(a)-C$ for A and P, respectively. In the no investigation case A receives s_w again, but P does not know the real outcome but knows only x_m . Therefore P gets x_m-s_w , and A collects the difference between x and x_m as well, such that he gets s_w+x-x_m .

Proposition 4: Assume C so small that (7) holds, communication and $\underline{s}_l < \underline{s}$. Then the optimal investigation strategy is random and the probability of investigation $\alpha(x)$ decreases in x for almost every x .

Proof:

The MSC is of the following form:

$$U(s_w) \geq \max_{x_m} \{\alpha(x_m)U(\underline{s}_l) + (1-\alpha(x_m))U(s_w+x-x_m)\} \quad \text{for all } x$$

or, since the RHS strictly increases in x , it must hold that

$$U(s_w) \geq \max_{x_m} \{\alpha(x_m)U(\underline{s}_l) + (1-\alpha(x_m))U(s_w+\bar{x}-x_m)\} .$$

Therefore $\alpha(x_m)$ must be chosen to guarantee this. Relabelling $x_m = x$ this gives

$$U(s_w) \geq \alpha(x)U(\underline{s}_l) + (1-\alpha(x))U(s_w+\bar{x}-x)\} \quad \text{every } x$$

or

$$\alpha(x) \geq \frac{U(s_w+\bar{x}-x)-U(s_w)}{U(s_w+\bar{x}-x)-U(\underline{s}_l)} . \tag{16}$$

But changing α from $\alpha_w = 1$ to $\alpha(x)$ affects also the ASC.

$$U(s_w) - V(a_w) \geq \max_{a'} \{U(\underline{s}_l)\int \alpha(x)f(x|a')dx + U(s_w)\int (1-\alpha(x))f(x|a')dx - V(a')\} =$$

$$= U(s_w) + \max_{a'} \{-[U(s_w) - U(\underline{s}_l)]\int \alpha(x)f(x|a')dx - V(a')\} .$$

For this to hold it must hold that

$$V(a_w) \leq [U(s_w) - U(\underline{s}_l)]\int \alpha(x)f(x|a)dx + V(a) \qquad \text{for all } a$$

or

$$\int \alpha(x)f(x|a)dx \geq \frac{V(a_w)-V(a)}{U(s_w)-U(\underline{s}_l)} \qquad \text{for all } a . \tag{17}$$

With these constraints the expected utility of P is

$$\int xf(x|a_w)dx - s_w - C\int \alpha(x)f(x|a_w)dx \tag{18}$$

from which it is clear that in order to maximize (18) P must minimize expected $\alpha(x)$. But both the constraints (16) and (17) must be satisfied. Denote the lowest value satisfying (16) by $\alpha_2(x)$, and (17) by $\alpha_1(x)$, respectively. Then the optimal $\alpha(x) = \max\{\alpha_1(x), \alpha_2(x)\}$.

To prove the assertion that $\alpha(x)$ decreases we derive some properties of $\alpha_1(x)$ in (17) and $\alpha_2(x)$ in (16). Now (16) is the stricter restriction on $\alpha(x)$ in the sense that it defines minimum values of $\alpha_2(x)$ for every x , whereas $\alpha_1(x)$ is determined only implicitly by the averaging condition (17).

From (16) it easily follows that $\alpha_2(x) < 1$ for all x since

$$U(s_w+\bar{x}-x) \geq U(s_w) > U(\underline{s}_l) .$$

The RHS of (16) is minimal for $x = \bar{x}$ for which $\alpha_2(\bar{x}) = 0$. By taking the first derivative of RHS $\alpha_2'(x) < 0$ for all x is easily verified.

Next consider $\alpha_1(x)$ as defined in (17).

The highest value of the RHS of (17) is $\frac{V(a_w)-V(\underline{a})}{U(s_w)-U(\underline{s}_l)}$, which is greater 0 (since $V(a_w)-V(\underline{a}) > 0$, and $U(s_w)-U(\underline{s}_l) > 0$), and is less than 1 ($V(a_w)-V(\underline{a}) < U(s_w)-U(\underline{s}_l)$, which follows by reformulating into $U(\underline{s}_l)-V(\underline{a}) < H_m = U(s_w)-V(a_w)$, and the inequality follows from $\underline{s}_l$ sufficient).

Consider the case if

$$\int \alpha_2(x) f(x|a)dx \geq \frac{V(a_w)-V(a)}{U(s_w)-U(\underline{s}_l)} \tag{19}$$

holds for all a, $0 \leq a \leq a_w$. Then $\alpha(x) = \alpha_2(x)$ for all x, and we conclude $\alpha'(x) < 0$ for all x.

If (19) does not hold for some a, $\alpha(x)$ must be increased for some x such that (17) will hold true. The subproblem is

$$\min \int \alpha_1(x) f(x|a_w)dx \tag{20}$$

subject to (17). It cannot be optimal to have

$$\int \alpha_1(x) f_a(x|a)dx \geq 0 \qquad \text{for all } a \tag{21}$$

since then (17) holds as an equality for $\underline{a}$. Now the RHS of (17) strictly decreases in a. But the LHS increasing contradicts the goal of minimizing (20). Therefore $\alpha_1{}'(x) > 0$ for all x, and $\alpha_1{}'(x) = 0$ for all x are not optimal, since (21) would hold (for $f_a(x|a)$ increases in x).

To get more specific, we follow an approach taken in Baiman and Demski [1980b] and Dye [1986]. Consider the pointwise Lagrangian corresponding to the subproblem (20) and (17).

$$L(x) = \alpha_1(x) f(x|a_w) - \int_{\underline{a}}^{a_w} \mu(a)[\alpha_1(x) - \frac{V(a_w)-V(a)}{U(s_w)-U(\underline{s}_l)}] f(x|a)da .$$

Taking the derivative with regard to α_1 gives

$$K(x) \equiv f(x|a_w) - \int_{\underline{a}}^{a_w} \mu(a) f(x|a)da .$$

$K(x) = 0$ can hold only by coincidence. Therefore $K(x) > 0$ gives $\alpha_1(x) = 0$, and $K(x) < 0$ gives $\alpha_1(x) = 1$. We know that

$$0 < \int \alpha_1(x) f(x|a) dx < 1 \qquad \text{for all } a$$

which implies that there exist some x for which $K(x) > 0$, and some x for which $K(x) < 0$. Rewrite $K(x)$ as

$$K(x) = f(x|a_w)[1 - \int_{\underline{a}}^{a_w} \mu(a) \frac{f(x|a)}{f(x|a_w)} da] .$$

Now $f(x|\cdot) > 0$ by assumption, and $\mu(a) \geq 0$. By MLRP $f(x|a_w)/f(x|a)$ increases in x for $a < a_w$, therefore the term in square brackets increases in x. This means that $K(x)$ must be negative for a lower subset of outcomes, and positive for the complementary subset. Hence $\alpha_1(x) = 1$ for $x \in [\underline{x}, x')$, and $\alpha_1(x) = 0$ for $x \in (x', \bar{x}]$.

Collecting these results both the constraints give weakly decreasing probabilities, therefore $\alpha(x)$ must be weakly decreasing, as well.

Q.E.D.

First note that Proposition 4 implies that there are *strict* gains to communication. It is often recognized that the available penalty must be bounded from below since otherwise combined with extremely low probabilities of investigation would make the problem under consideration trivial. The reason is that the first-best solution could be approached arbitrarily closely (as expected cost of investigation approach zero). Here we get the result that for any sufficient penalty given exogenously the optimal probability of investigation decreases in x.

The result confirms the conjecture by Townsend [1979] that random investigation with decreasing probability is Pareto-superior to the interval investigation contract he considered. It was also shown for the case of pure risk-sharing by Mookherjee and Png [1986] where it arises from a condition similar to (16). They extend their analysis to the case of moral hazard but do not assume observability of the action in case of investigation. Therefore they do not get our condition (17) which gives weakly decreasing probabilities. On the other hand their result goes beyond ours just presented since they do not restrict to the case of sufficient penalty as we did in this section. We will pursue the case of insufficient penalty in the following section.

Observe, as well, that condition (17) is comparable to the ASC in the model of Kanodia [1985]. He derives decreasing $\alpha(x)$ by assuming x to be public information, but A having predecision information about the environment which enables deleting the integral in front of $\alpha(x)$ in (17). Of course this gives strictly decreasing $\alpha(x)$ since the RHS strictly decreases. But his result comes from the predecision information assumption since only assuming x public information (i.e. deleting the MSC in the model) gives optimality of a lower interval investigation (Dye [1986]).

In the characterization of Proposition 4 the cost of investigation was assumed to be bounded from above. The reason for this restriction lies in the fact that - according to Proposition 1 - otherwise the PRC dominates the PWC. But observe that in case of communication the situation changes as the PWC actually can be achieved without fully investigating. As noted above assuming very low penalty the probability of investigation approaches a value near zero. This has the effect that even if C is high, combined with a very low α the importance of C diminishes. Therefore even if C is such that (7) does not hold there exists a set of $\underline{s}_I \in (-\infty, \underline{s}'')$, $\underline{s}'' < \underline{s}$, such that the RIC is strictly preferable to the PRC. This states a trade-off between cost C and sufficient penalty $\underline{s}_I$. The higher C the lower $\underline{s}_I$ must be to achieve an improvement on the PRC. Observe that if the RIC is preferable the results on the investigation strategy in Proposition 4 will continue to hold.

6. Communication and Insufficient Penalty

Recalling the program (12) - (15) it is obvious that it is harder to solve for insufficient penalty than for sufficient penalty. This is because we do not know the compensation scheme $s(x)$. Therefore the next result actually is a corollary to Proposition 4 for it provides a partial extension to insufficient penalty. For the characterization of the investigation strategy we will assume some properties of the compensation scheme.

Corollary: Assume C *so small that (7) holds,* $\underline{s}_h > \underline{s}$ *. Assume further*

(i) $0 \le s'(x) \le 1$ *for all* x *,*

(ii) $s(\underline{x}) \ge \underline{s}_h$ *.*

Then the optimal contract is random with probabilities of investigation $\alpha(x) > 0$ *for all* $x < \bar{x}$ *, and* $\alpha(x) \ge \alpha_3(x)$ *for all* x *, where* $\alpha_3'(x) < 0$ *.*

Proof:

Under assumptions (i) and (ii) the MSC (15) is satisfied by

$$\alpha(x) \ge \alpha_3(x) = \frac{U(s(x)+\bar{x}-x)-U(s(\bar{x}))}{U(s(x)+\bar{x}-x)-U(\underline{s}_h)}$$

since (i) ensures the RHS increasing in x , and (15) holding for every x , it holds for $\bar{x}$. Then $\alpha_3(x)$ has the same properties as for sufficient penalty (see Proposition 4), viz. $0 \le \alpha_3(x) \le 1$, $\alpha_3(x) > 0$ for all $x < \bar{x}$, $\alpha_3(\bar{x}) = 0$, and $\alpha_3'(x) < 1$ for all x .

Q.E.D.

This corollary is a partial characterization of the investigation strategy for insufficient penalty. It follows from considering the impact of the MSC on the optimal probability of investigation under some assumptions regarding properties of the compensation scheme. The optimal probability of investigation is strictly greater 0 for all outcomes except the highest outcome (for which the corollary includes no statement). And the probabilities are at least as great as a probability function with the property that it is strictly decreasing.

Assumption (i) on the compensation scheme is not very strong. It states that $s'(x) \ge 0$ which is intuitively appealing because if $s'(x) < 0$ then A would prefer to choose the lowest feasible action since that increases his expected utility in the no investigation case. Furthermore it would provide no incentive to take the agreed action. The second part of (i) says that $s'(x) \le 1$ should hold.

Otherwise EG would decrease for increasing action at least for some a . Assumption (ii) is more severe, it follows from the MSC to be met. Observe that (15) could never be satisfied for $\alpha < 1$ if $s(x) < \underline{s}_h$ for some x and some x_m given assumption (i).

An example for decreasing probabilities to be preferable was shown by Fellingham [1980] by employing the MSC. The result is also similar to that of Mookherjee and Png [1986].

A condition for $\alpha(x)$ from the ASC like that for sufficient penalty in equation (17) turns out to be rather messy, and seems to give no direct result for $\alpha(x)$ in general. The same is true for the optimal compensation scheme $s(x)$, properties of which were assumed in the corollary.

The *example* introduced in section 3 will now be continued to show some results for sufficient penalty. For any $\underline{s}_l < \underline{s}$ this means to find a function $\alpha(x)$ which minimizes

$$\int \max \{\alpha_1(x), \alpha_2(x)\} f(x|a_w)dx ,$$

where $\alpha_1(x)$ and $\alpha_2(x)$ are defined in equations (16) and (17). Condition (16) implies

$$\alpha_2(x) = \frac{U(s_w+\bar{x}-x)-U(s_w)}{U(s_w+\bar{x}-x)-U(\underline{s}_l)} = \frac{U(\infty)-U(s_w)}{U(\infty)-U(\underline{s}_l)} = \frac{U(s_w)}{U(\underline{s}_l)} \quad \text{for } x < \infty \text{ .} \qquad (16')$$

This constant is the limiting degenerate function of a strictly decreasing function if the support of x is bounded from above. The optimal $\alpha(x) \geq \alpha_2$ here.

Condition (17) applied to the assumptions of the example gives

$$\int \alpha_1(x)e^{-x/a}dx \geq \frac{a(a_w-a)}{U(s_w)-U(\underline{s}_l)} \qquad \text{for all } a \text{ .} \qquad (17')$$

This holds for many (weakly) decreasing functions $\alpha_1(\cdot)$. For example

$$\alpha_1(x) = \max_a e^{-\frac{x}{a} \cdot \left[\frac{U(s_w)-U(\underline{s}_l)}{a_w - a} - 1\right]} \tag{22}$$

which is strictly decreasing, and $\alpha_1(\underline{x}) = 1$, $\alpha_1(\bar{x}) = 0$.

The function (22) must hold for all a , $0 < a \leq a_w$. After some calculation the necessary a' which guarantees that (22) will hold for all a turns out to be

$$a' = a_w - U(s_w) + U(\underline{s}_l) + [U(s_w)-U(\underline{s}_l)]\cdot[U(s_w)-U(\underline{s}_l)-a_w]^{1/2} .$$

This gives

$$\alpha_1(x) = e^{-\frac{x}{a'} \cdot \left[\frac{U(s_w)-U(\underline{s}_l)}{a_w - a'} - 1\right]} \tag{22'}$$

Another family of solutions to (17') is

$$\alpha_1(x) = \begin{cases} d & \text{for } x \leq \max_a -a\cdot\ln\left(1 - \frac{a_w - a}{d[U(s_w)-U(\underline{s}_l)]}\right) \\ 0 & \text{else} \end{cases} \tag{23}$$

for $\frac{a_w - a}{U(s_w)-U(\underline{s}_l)} < d \leq 1$.

A limiting result of (23) is the constant function over the full support of x

$$\alpha_1(x) = \frac{a_w - \underline{a}}{U(s_w)-U(\underline{s}_l)} = \text{constant for all } x . \tag{23'}$$

But from Proposition 4 it is clear that this cannot be optimal. A preferable function is now constructed by utilizing the constant α_2 (in this example).

$$\alpha_1(x) = \begin{cases} 1 & \text{for } x \leq \max_a -a\cdot\ln\left[\frac{1}{1-\alpha_2}\cdot\left(1-\frac{a_w - a}{d[U(s_w)-U(\underline{s}_l)]}\right)\right] \\ \alpha_2 & \text{else} \end{cases} \tag{24}$$

which shifts the mass of probability arising from the difference between α_1 and α_2 to the left of the support of x .

For illustration purposes I considered the solutions to (22'), (23'), and (24). Minimizing

$$\int \alpha(x) f(x|a_w) dx$$

gives strict preferability of (24) for a special case by tedious calculation.[7] Recalling the steps of the proof of Proposition 4 this seems to be the optimal solution.

Nevertheless, for drawing Figure 4 a constant α_1 as defined in (23') is assumed for computational convenience. Note that this is not the optimal investigation strategy, but is preferable to full investigation in case of sufficient penalty. With this assumption the expected utility of P conditional on the penalty available is depicted in Figure 4, summarizing results for both sufficient and insufficient penalty.

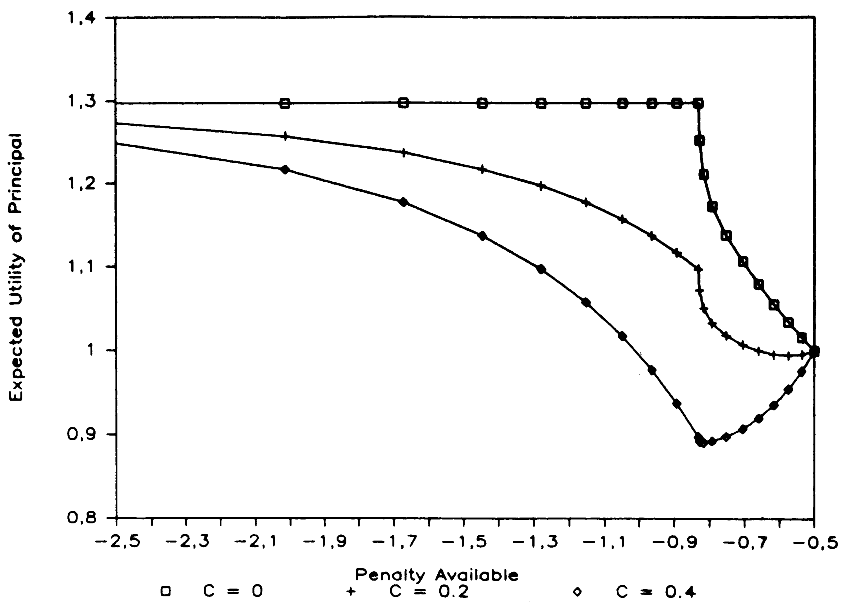


Figure 4: Expected utility of principal contingent on penalty available

[7] The results of this calculation, assuming $\underline{s}_1 = -2$, are as follows. The function defined in (22') gives $\alpha_1(x) = \exp(-13.597x)$. (23') gives $\alpha_1(x) = 0.2031$. (24) gives $\alpha_1(x) = 1$ if $x \leq 0.00864$, and 0.1353 else.

The preferability of (24) follows from calculating the difference of the integrals of $\alpha_1(x)$ and $\alpha(x)$. This gives 0.0312 for (22'), 0.0677 for (23'), and 0.0057 for (24), from which it is clear that (24) is preferable since it minimizes expected cost.

Additional information about the probabilities of investigation producing the results depicted can be visualized by Figure 4. The points indicated by symbols show the values of α , starting from the right with $\alpha = 0$, 0.1 , 0.2 , and so on, $\alpha = 1$ at $\underline{s}$, then decrease again in steps of tenths.

7. Conclusions

The purpose of this paper was to find conditions for random investigation to be optimal. In a rather "black and white" setting it has been shown that in a model without possibility of communication between the principal and the agent random investigation depends upon the availability of penalty that can be imposed upon the agent after detecting some deviation. The cost of investigation is not the crucial parameter for determining the probability of investigation. Next, it has been shown that in the case of allowing for communication where the agent tells the principal his superior information a much richer setting for non-trivial investigation strategies is found. The analysis was divided in the cases sufficient and insufficient penalty, respectively. Communication allows for utilizing sufficient penalty in a superior way than was possible without communication. Its availability prevents the agent from choosing any action or message the principal dislikes even though an investigation is not always conducted. Hence this saves expected investigation cost without changing other parameters. The optimal probability of investigation was found to be decreasing with higher outcomes. In case of insufficient penalty available the probabilities are strictly greater zero for all reported outcomes except the highest outcome, and are greater than or equal to a strictly decreasing investigation strategy. This last result is rather weak compared to that for sufficient penalty. It follows from the fact that the attempt to solve the program implicitly for all parameters is a quite messy task. Some properties of the compensation scheme were assumed, instead.

Most of the limits of the results follow from the simple setting. It is obvious that the results strongly depend on the assumption of the information set availability to the principal before deciding if to investigate, and to the agent before choosing his action. It would be interesting to know how the results change if some other costly investigation technology instead of independent perfect information is used, and what if costs of investigation vary. Presumably multiperiod considerations (e.g., reputation) will alter the results as well.

References

BAIMAN, S. "Discussion of Auditing: Incentives and Truthful Reporting." *Journal of Accounting Research* (Supplement 1979): 25 - 29.

BAIMAN, S., AND J. S. DEMSKI "Variance Analysis Procedures as Motivational Devices." *Management Science* (August 1980a): 840 - 848.

BAIMAN, S., AND J. S. DEMSKI "Economically Optimal Performance Evaluation and Control Systems." *Journal of Accounting Research* (Supplement 1980b): 184 - 220.

DEMSKI, J. S., AND G. A. FELTHAM "Economic Incentives in Budgetary Control Systems." *Accounting Review* (April 1978): 336 - 359.

DYE, R. A. "Optimal Monitoring Policies in Agencies." *Rand Journal of Economics* (Autumn 1986): 339 - 350.

EVANS, J. H. "Optimal Contracts with Costly Conditional Auditing." *Journal of Accounting Research* (Supplement 1980): 108 - 128.

FELLINGHAM, J. C. "Discussion of Optimal Contracts with Costly Conditional Auditing." *Journal of Accounting Research* (Supplement 1980): 129 - 133.

HOLMSTRÖM, B. "Moral Hazard and Observability." *Bell Journal of Economics* (Spring 1979): 74 - 91.

KANODIA, C. S. "Stochastic Monitoring and Moral Hazard." *Journal of Accounting Research* (Spring 1985): 175 - 193.

LAMBERT, R. A. "Variance Investigation in Agency Settings." *Journal of Accounting Research* (Autumn 1985): 633 - 647.

MOOKHERJEE, D., AND I. PNG "Optimal Auditing, Insurance, and Redistribution." Working Paper (Stanford and UCLA), June 1986.

MYERSON, R. B. "Incentive Compatibility and the Bargaining Problem." *Econometrica* (January 1979): 61 - 73.

ROGERSON, W. P. "The First-Order Approach to Principal-Agent Problems." *Econometrica* (November 1985): 1357 - 1367.

SAPPINGTON, D. "Limited Liability Contracts between Principal and Agent." *Journal of Economic Theory* 29 (1983): 1 - 21.

TOWNSEND, R. M. "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory* 21 (1979): 265 - 293.

YOUNG, R. A. "A Note on 'Economically Optimal Performance Evaluation and Control Systems': The Optimality of Two-Tailed Investigation." *Journal of Accounting Research* (Spring 1986): 231 - 240.

Section 6 Coordination in Groups

Managers as Principals and Agents

Martin J. Beckmann

Summary: The principal owns a simple organization in which an agent supervises operatives. The agent chooses his/her own effort as well as the number of operatives to be hired. Operatives receive fixed wages and the agent a share of profits. In this model explicit solutions are given for the agent's chosen effort and the agent's optimal profit share. Increasing returns to scale lead to richer results than the constant returns to scale case.

Table of Contents

1. Introduction

We consider the following model: The absentee owner (principal) of a simple organization hires a manager (agent) to run the organization. The manager hires and supervises x_0 operatives, who must work full time, but the manager is free to choose his/her own level of effort x_1. Output y is a function of the two inputs x_0, x_1,

$$y = F(x_0, x_1) .$$

Operatives receive a unit wage $w_0 = 1$. The manager receives a share δ of profits, and the effort level x_1 is a function of this share δ. Which share δ maximizes the owner's profits?

This model is mathematically and conceptually simple, but we believe relevant. In this simple formulation it is intended mainly as a con-

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

tribution to Organization Theory.

2. Managers as Principals; Span of Control

We begin by modelling a production function for the manager's output. Suppose that the main task of management is to monitor the performance of operatives. An operative is expected to work full time $t = 1$ but may choose to shirk, be idle, part of the time. Let an operative's utility function be

$$u = h \cdot \ell n(2 - t) + w_o \tag{1}$$

where

u utility

w_o income

$h > 0$ a parameter measuring the attractiveness of leisure.

When $t < 1$, the operative is idle for the fraction $1 - t$ of working time. If $(1 - t)p$ is the probability of being caught and k the penalty imposed in the form of lost wages or reprimands then the objective of an operative is assumed to be the maximization of utility in terms of leisure and expected income,

$$\max_{0 \le t \le 1} h \ell n(2 - t) + (1 - t)p(-k) + w_o \, .$$

A solution $t < 1$ is characterized by

$$t = 2 - \frac{h}{k}\frac{1}{p} < 1 \, . \tag{2}$$

Let now p be the proportion of time that the supervisor allocates to the x_o operatives. If this allocation is random then the probability p of being caught shirking is

$$p = \frac{x_1}{x_o} \, ,$$

yielding

$$t = 2 - \frac{h}{k}\frac{x_o}{x_1} \, . \tag{3}$$

To fix the parameter h consider how much a person would work voluntarily for a unit reward on his/her own:

$$\max_{t} h \cdot \ell n(2 - t) + t \, .$$

This is solved by $t = 2 - h$. If this working time is assumed to be unity then $h = 1$. Substituting in (3) yields

$$t = 2 - \frac{1}{k}\frac{x_o}{x_1} \, . \tag{4}$$

Equation (4) shows that there is a ratio of operatives to managers or span of control $\frac{x_0}{x_1}$ that involves sufficient supervision per operative to guarantee full time work $t = 1$,

$$\frac{x_0}{x_1} = k \tag{5}$$

An alternative interpretation for $\frac{x_0}{x_1}$ is: ratio of intended working time by operatives to control time by supervisor.

It turns out that this span of control equals the penalty for shirking. The output y produced (by manager's effort x_1) is now assumed to be proportional to x_0,

$$x_0 = \frac{1}{b} y$$

or, using (5)

$$x_1 = \frac{1}{bk} y .$$

Output is thus proportional to x_1. The organization operates with fixed coefficients. We have a Leontief production function. The profit g equals output y minus the wage bill x_0,

$$g = y - x_0 = (b-1)x_0 ,$$

or

$$g(x_1) = (b-1)k\, x_1 . \tag{6}$$

Thus profit is proportional to managerial effort x_1.

3. The Production Function

The combination of managerial and operative inputs in an organization need not take the specific form described so far. It will however be subject to some type of production function $F(x_0,x_1)$ describing the maximal output an organization can achieve when a manager puts in x_1 time units of effort and the operatives supply x_0 time units of effort. (For a general description cf. Beckmann, 1983).

3.1 Linear Homogeneous Production Function

That profit is proportional to managerial effort x_1 is true for all linear homogeneous production functions. This may be seen as follows:

$$g(x_1) = \max_{x_0} \{F(x_0,x_1) - x_0\} = x_1 \max_{x_0} \{\hat{F}(\frac{x_0}{x_1}) - \frac{x_0}{x_1}\} = x_1 g_1 \tag{7}$$

where

$$g_1 := \max_{z} \{\hat{F}(z) - z\}, \hat{F}(z) = F(z,1) .$$

3.2 Cobb Douglas Production Function

As an example we may consider the linear homogeneous Cobb Douglas production function,

$$y = bx_0^{\alpha} x_1^{\beta} \quad \text{with} \quad \alpha = \frac{3}{4}\,,\ \beta = \frac{1}{4}\ . \tag{8}$$

With this Cobb Douglas production function, profits as a function of managerial effort are

$$g(x_1) = \underset{x_0}{\text{Max}}\ \{bx_0^{\alpha} x_1^{\beta} - x_0\} = x_1\ \underset{x_0}{\text{Max}}\ \{b(\frac{x_0}{x_1})^{\alpha} - \frac{x_0}{x_1}\} = x_1 \cdot g_1$$

where g_1 is determined by

$$g_1 = (1-\alpha)b^{\frac{1}{1-\alpha}}\alpha^{\frac{\alpha}{1-\alpha}}\ . \tag{9}$$

To determine the level factor b assume that the product or service produced by the organization can also be supplied by an individual operating on his/her own under the same production function. If a single person allocates his/her time among management and operative labour , $x_0 + x_1 = 1$, then an optimal allocation of effort requires

$$\frac{\partial y}{\partial x_0} = \frac{\partial y}{\partial x_1}\ ,$$

or

$$\frac{\alpha y}{x_0} = \frac{\beta y}{x_1}\ .$$

It follows

$$\frac{x_0}{x_1} = \frac{\alpha}{\beta}$$

or
$$y = b\alpha^{\alpha}\beta^{\beta} = g\ .$$

A unit level of effort achieves profits g if and only if

$$b = g\alpha^{-\alpha}\beta^{-\beta}\ .$$

Specifically, if an independent individual earns 50% more than a hired operative, then

$$b = 1.5\ (\frac{3}{4})^{-3/4}\ (\frac{1}{4})^{-1/4} = 2.6178\ .$$

From this using (9)

$$g_1 = \frac{1}{4}\ (2.6178)^4 \cdot (\frac{3}{4})^3 = 4.955\ .$$

We shall compare the "marginal productivity" g_1 of a manager to that of an operative. Given k (= span of control) the profit is

$$g_k(x_0) = F(x_0, \frac{x_0}{k}) - x_0 = b\ x_0^{\alpha}(\frac{x_0}{k})^{1-\alpha} - x_0 = b(b\cdot k^{\alpha-1} - 1)x_0\ .$$

The marginal productivity

$$g'_k(x_0) = b \cdot k^{\alpha-1} - 1$$

is independent of the number x_0 of operatives. For our example and a span of control $k = 6$ we have

$$g'_6(x_0) = 2.6178 \cdot 6^{-0.25} - 1 = 0.6726 \ .$$

Hence the marginal productivity (4.955) of a manager is more than seven times that of an operative.

4. Managers as Agents

Next we determine the level of effort chosen by the manager as a function of his/her rewards. The reward may take any form: a fixed salary w_1 or a share δ of total profits g are two possibilities. Between these extremes may be found other types of compensation usually consisting of some fixed payment and some bonus that is proportional to achieved profits.

4.1 Manager's Effort

Here we shall consider that he receives only a share δ of total profits g. When the production function is linear homogeneous this reward turns out to be proportional to managerial effort $\delta g = \delta g_1 x_1$. To determine the manager's voluntary effort one must consider his/her utility function in terms of leisure $2 - x_1$, and money income $\delta g_1 x_1$. Effort is measured in time units and total available time is 2 units. As before let utility be additive, logarithmic in terms of leisure, and linear in terms of income,

$$u = h \, \ell n(2 - x_1) + \delta g_1 x_1 \ . \qquad (10)$$

The factor h measures the attractiveness of leisure in relation to that of income and may be different for managers and operatives. A utility maximizing manager chooses an effort level x_1 such that

$$0 = \frac{du}{dx_1} = \frac{-h}{2-x_1} + \delta g_1$$

from which follows

$$x_1 = 2 - \frac{h}{\delta g_1} \ . \qquad (11)$$

This is an increasing function of share δ and may be considered the manager's supply function of effort (Fig. 1).

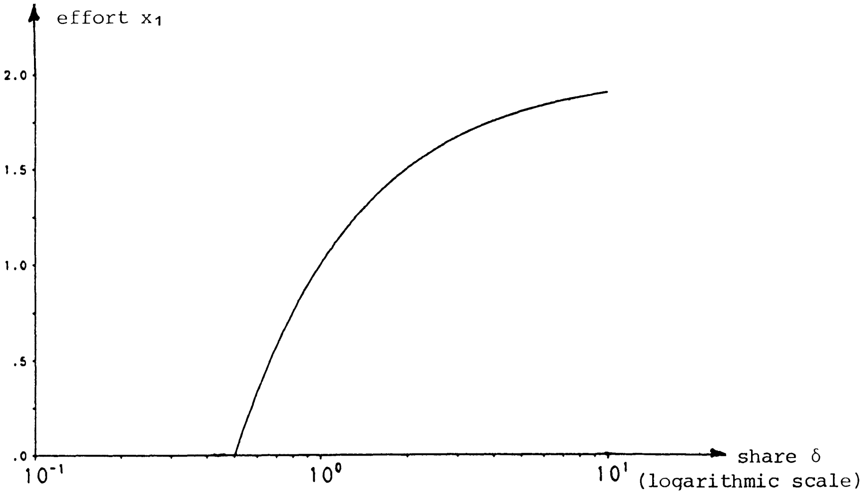


Fig. 1: Manager's effort as function of profit share ($h = g_1 = 1$)

The manager's achieved income is then $\delta g_1 x_1 = 2\delta g_1 - h$. This is positive provided the profitability g_1 of the organization and the manager's share δ are large enough relative to the strength h of the preference for leisure. When $h = 1$ (normal preference for leisure) then the manager's income exceeds that of a worker $w_0 = 1$ provided δg_1 exceeds unity.

4.2 Optimal Share; Comparison with v.Thünen's Wage Formula

Consider now the owner's income. It is $z = (1 - \delta)g$ or, in view of (7),(11),

$$z = (1 - \delta)g_1 x_1 = 2g_1(1 - \delta) - \frac{h(1-\delta)}{\delta} . \tag{12}$$

Maximization of owner's income with respect to the manager's share δ yields

$$0 = \frac{\partial z}{\partial \delta} = -2g_1 - \frac{-h\delta - h(1-\delta)}{\delta^2}$$

or

$$\delta = \sqrt{\frac{h}{2g_1}} . \tag{13}$$

Using (13), the manager's reward in terms of his/her input x_1 and productivity g_1 becomes

$$\delta g = \sqrt{\frac{h}{2}\frac{g^2}{g_1}} = x_1 \sqrt{\frac{h}{2} g_1} . \tag{14}$$

This is reminiscent of von Thünen's famous wage formula: an employee's compensation is to be set proportional to the square root of his/her productivity g_1 . The term $\frac{h}{2}$ would have to represent subsistence income for a full analogy.

Suppose subsistence income s is defined as that which at full time work yields just enough utility to make a person indifferent to not working at all; then

$$h \ln 1 + s = h \ln 2$$

or

$$s = h \ln 2$$
$$= 0.693\ h > \frac{h}{2} \ .$$

Except for a small numerical difference (between $\frac{1}{2}$ and 0.693) von Thünen's formula applies throughout.

4.3 Cost of Agency

Suppose a manager as owner puts in a full time effort $x_1 = 1$. This means that a maximum of $g_1 x_1 + h \ln (2 - x_1)$ is achieved when $x_1 = 1$. Necessary and sufficient for this is that $h = g_1 (= g)$. If these values are used as benchmarks and substituted in the share formula (13) one obtains

$$\delta = \sqrt{\frac{1}{2}} = 0.7071$$

as the optimal manager's share, i.e. that share which maximizes owner's return under the restriction $x_1 = 1$. The owner's income is then (compare (12) and recall $h = g_1 = g$)

$$z^* = g \cdot (2 - \sqrt{2})(1 - \frac{1}{\sqrt{2}}) = 0.1716\, g \qquad (15)$$

compared to g which the owner can earn on his own. Notice that this is valid for all linear homogeneous production functions. The cost of agency thus turns out to be 83%.

4.4 Multi-Level Organization

So far we considered simple organizations, requiring only one managerial level. In multi-level organizations an agency problem could arise at every level. Suppose however that it is only top management - the president - who can freely choose his/her level of managerial effort, while all lower ranking managers are fully supervised and require no additional incentive. Now the managerial mode of "management by delegation" implies in fact constant returns to scale, i.e. a linear homogeneous production function for the entire organization (Beckmann 1983, pp 151-159). The previous analysis is thus applicable to the compensation of top mangement in hierarchical organizations.

5. Increasing Returns to Scale

In simple organizations we may have increasing returns to scale. Let $F(x_0, x_1)$ be homogeneous of degree $m > 1$ and Cobb Douglas. Now, for given x_1 (and $\alpha < 1$),

$$g(x_1) = \underset{x_0}{\mathrm{Max}}\ b x_0^{\alpha} x_1^{\beta} - x_0$$

is achieved for $\alpha b x_1^{\alpha-1} x^{\beta} = 1$ or $x_0 = (\alpha b x_1^{\beta})^{\frac{1}{1-\alpha}}$.

Substituting in (16) yields

$$g(x_1) = (1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}} b^{\frac{1}{1-\alpha}} x_1^{\frac{\beta}{1-\alpha}} = g_1 x_1^{\frac{\beta}{1-\alpha}} \quad . \tag{17}$$

Notice that $\beta/(1-\alpha) > 1$, in view of $m = \alpha + \beta > 1$.

5.1 Manager's Effort

The manager's earnings as a function of effort x_1 are

$$\delta \cdot g_1 \cdot x_1^{\frac{\beta}{1-\alpha}} \tag{18}$$

and his/her utility function is

$$u = h\, \ell n(2 - x_1) + \delta g_1 \cdot x_1^{\frac{\beta}{1-\alpha}} \quad .$$

The first term is concave, the second convex. Maximization with respect to $x_1 \geq 0$ yields in general two local maxima which must be compared

$$x_1 = 0 \quad \text{if} \quad \left.\frac{\partial u}{\partial x_1}\right|_0 \leq 0 \quad \text{(corner solution)} \tag{19}$$

$$0 = \frac{\partial u}{\partial x_1} \quad \text{for} \quad x_1 = \hat{x}_1 > 0 \quad \text{provided} \quad \left.\frac{\partial^2 u}{\partial x_1^2}\right|_{\hat{x}_1} < 0 \ . \tag{20}$$

Condition (19) is always satisfied since $\left.\frac{\partial u}{\partial x_1}\right|_0 = -\frac{h}{2} < 0$.
Condition (20) states that

$$0 = -\frac{h}{2-\hat{x}_1} + \delta g_1 \frac{\beta}{1-\alpha} \hat{x}_1^{\frac{\alpha+\beta-1}{1-\alpha}} \tag{21}$$

provided the second order condition is satisfied,

$$-\frac{h}{(2-\hat{x}_1)^2} + \delta g_1 \frac{\beta}{1-\alpha} \cdot \frac{\alpha+\beta-1}{1-\alpha} \hat{x}_1^{\frac{\alpha+\beta-1}{1-\alpha}-1} < 0 \ .$$

Suppose for instance that

$$\alpha = \frac{3}{4}, \quad \beta = \frac{1}{2} \ ,$$

then (21) becomes $(2-\hat{x}_1)\hat{x}_1 = a$, where

$$a = \frac{h}{2\delta g_1} \tag{22}$$

The solution satisfying the second order condition is

$$x_1 = 1 + \sqrt{1 - a} \tag{23}$$

It dominates the corner solution $x_1 = 0$ when

$$\ln(1 - \sqrt{1 - a}) + \frac{(1 + \sqrt{1 - a})^2}{2a} > \ln 2 \tag{24}$$

In terms of $\rho = \sqrt{1 - a}$

$$\ln(1 - \rho) + \frac{(1 + \rho)^2}{2(1 - \rho^2)} > \ln 2 \qquad \text{or}$$

$$\ln(1 - \rho) + \frac{1}{1 - \rho} > \frac{1}{2} + \ln 2 = 1.1931472 \tag{25}$$

$$\varphi - \ln \varphi > \frac{1}{2} + \ln 2 \qquad \varphi = \frac{1}{1 - \rho}$$

which is true for $\varphi > 1.7564$ or $a < 0.8145$

i.e. when

$$\frac{\delta g_1}{h} > \frac{2}{0.8145} = 2.4555 \ . \tag{26}$$

Otherwise the agent prefers to do nothing.

5.2 Feasibility Conditions and Optimal Share

The owner's earnings are

$$z = (1 - \delta)g(x_1) = (1 - \delta)g_1 \left\{1 + \sqrt{1 - \frac{1}{2\delta}\frac{h}{g_1}}\right\}^2 \ . \tag{27}$$

Maximizing z with respect to δ yields

$$0 = -(1 + \sqrt{1 - \tfrac{c}{\delta}})^2 + (1 - \delta)(1 + \sqrt{1 - \tfrac{c}{\delta}}) \frac{1}{\sqrt{1 - \frac{c}{\delta}}} \cdot \frac{c}{\delta^2}$$

where $c = \dfrac{h}{2g_1}$. (28)

Straightforward manipulations reduce this to a cubic in δ

$$\delta^3 - 2\delta^2 + c = 0 \ . \tag{29}$$

Implicit differentiation shows that

$$\frac{d\delta}{dc} = - \frac{1}{3\delta^2 - 4\delta} > 0 \text{ for all } \delta \leq 1 \ . \tag{30}$$

This means that the optimizing share δ increases with c and this in turn states that the optimal managerial share rises with preference for leisure h and decreases with productivity g_1 . A half share results when

$$c = \frac{3}{8}$$

which would require a small preference for leisure or high productivity. Thus for $h = 1$ a productivity rate $g_1 = \frac{4}{3}$ would be required.

The manager then puts in (an almost incredible) effort

$$\hat{x}_1 = 1 + \sqrt{1 - \frac{3}{4}} = \frac{3}{2} .$$

The resulting profit is $g_1 \cdot \frac{9}{3} = 3$ and both manager and owner would receive $\frac{3}{2}$ units of profit resulting in manager's utility of

$$\ln \frac{1}{2} + \frac{3}{2} = 0.80685 .$$

It appears from this as compared to (15) that increasing returns to scale are beneficial to both principal and agent.

When the condition (26) is violated so that $a > 0.8145$ then the optimal solution is $\hat{x}_1 = 0$.

The agent refuses the contract. But this is not in the owner's interest. To avoid this the owner must agree to a share of

$$\delta > 2.4555 \frac{h}{g_1}$$

In turn this is practical only when $\delta < 1$ or

$$h < 1.629\, g_1 \tag{31}$$

The agent's preference for leisure must not exceed 1.629 times the productivity rate g_1. Inequality (31) states the condition under which agency is possible in the operation of a simple organization with (special) increasing returns to scale.

References

Beckmann, M.J. (1983): Tinbergen Lectures, Springer-Verlag, Berlin-Heidelberg-New York

Siglitz, J.E. (1974): Risk Sharing and Incentives in Sharecropping, Review of Economic Studies 61, 219-256.

Misperceptions, Equilibrium, and Incentives in Groups and Organizations

Martin Gaynor and Paul R. Kleindorfer

SUMMARY: This paper considers multi-agent principal-agent problems in an organizational or group context. The principal sets incentives for a group of agents who then adjust their behaviors to equilibrium either cooperatively or noncooperatively. For a fixed production technology, we state the principal's design problem and survey recent results concerning its solution. We then consider the effects of agents' misperceptions of the production technology on their equilibrium behavior (and consequently also on the principal's design problem). We define a consistent equilibrium as one at which, whatever misperceptions may be present, each agent receives in equilibrium the payment he or she expected on the basis of the agent's (possibly misperceived) production technology and promised incentives from the principal. We provide sufficient conditions under which such a consistent equilibrium exists. Several recent empirical studies on expectations are reviewed at the conclusion of the paper, the results of which are compatible with the consistent equilibrium properties under misperceptions which we study here.

TABLE OF CONTENTS

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

1. INTRODUCTION

The topic we investigate in this paper is the effect of misperceptions on the part of members of a group or an organization on the incentive mechanisms designed to induce optimal behavior by the members. This work is in the spirit of multiple-agent principal-agent problems. The principal here is viewed as an organization or group "designer" who chooses incentives and information structures so as to maximize his net residual benefits. The standard method for analyzing such problems is to treat the principal as a von Stackelberg leader, whose choice space is a set of possible (e.g. incentive) designs, with agents then playing a sub-game determined by the principal's choice of design. Our departure from most previous work is that we allow agents to have less than perfect perceptions of environmental parameters, such as incentives, technology, and other agents' behaviors. This bounded rationality assumption seems especially appropriate for organizational choice, where a wide body of research supports the existence of heuristics and biases in choice behavior. Moreover, the notion of misperceptions as a theoretical construct provides interesting implications for the principal-designer's organizational design problem. As a prelude, it will be useful to discuss the underlying themes from the economics or organization literature which motivated this study.

The literature on economics of organizations is usually traced to Coase's (1937) paper on the nature of the firm. The questions posed by Coase were the first look inside the "black box" constituting the theory of the firm. Coase makes clear that the standard theory of the firm is in actuality an arbitrary imposition, rather than a necessary consequence of basic economics. He focuses on the role of transactions costs and uncertainty. Coase's contention is that the firm serves to economize on transactions costs and it can best be analyzed in this way. This discussion by Coase raised a number of questions about why some activities are organized within firms rather than outside them. The pursuit of these questions has led to the study of the economics of organizations.

The matter was left to rest until the work of Simon, Cyert, and March at Carnegie-Mellon. Simon (1957), and Cyert and March (1963) applied psychological models of human behavior to the analysis of behavior in organizations. This led to the insights of bounded rationality and differing motives for the various actors within a firm or organization. Williamson (1964), building upon these ideas and those of Berle and Means (1932), and Gordon (1961) on the separation of ownership and control, incorporated them into an analysis of discretionary behavior on the part of managers. Both Williamson, and Marris (1964) added the influence of monopoly power into this analysis. Managerial discretion is enhanced by monopoly power, which gives managers more rents over which to have discretion. This hypothesis has been labeled "expense-preference theory" and is related to Leibenstein's (1966) independent work on organizational slack and x-inefficiency.

While Williamson and Leibenstein advanced the economic theory of organizations by investigating the phenomenon of "slack", an impossible occurrence in the neo-classical theory of the firm, neither of these efforts were directed at generating a general economic theory of organizations. A major step in generating such a theory occured in 1972 with the publication of The Economic Theory of Teams by Marschak and Radner and "Production, Information Costs and Economic Organization" by Alchian and Demsetz.

Team theory assumes a common preference function for all members of the team. Thus, conflict does not exist within a team, but there are problems of communication. The problem of the manager is to elicit true messages from his subordinates, as in the free-rider preference revelation problem in public finance[1].

Alchian and Demsetz (1972) focus on situations in which conflict of interest exists between members of an organization, either within or across hierarchical boundaries. The emphasis is on the transactions costs of organizing, monitoring, metering and enforcing contracts within the firm. The objectives of employer and employee differ, production is not separable in employees' inputs, and the

[1]It should be noted that the solution to the free-rider preference-revelation problem proposed by Groves and Ledyard (1977) is derived from Groves' (1973) work on team theory, which in turn is derived from previous work by Marschak and Radner, as summarized in their book.

employer cannot costlessly observe the behavior of the employees. This sort of situation is termed moral hazard in that once a contract of employment is granted the employee may exploit opportunities, unobservable to the employer, to act in ways which are not in the employer's best interest. Moral hazard is not present in traditional Marschak-Radner team theory, since all members of an organization possess identical objectives.

Contemporaneously with the raising of issues in the internal theory of the firm, the methodology of game theory was applied to the analysis of the incentive properties of bilateral contracts between a "principal" and his "agent" when there exists an asymmetry of information. Originally, the analysis of moral hazard was confined to analysis of bilateral exchanges, as exemplified by the agency theory of Ross (1973). More recently, agency theory has been extended to the case of many agents. This extension allows the application of this form of analysis to the economics of organizations. This extension is a natural one, since previous research (e.g., Alchian and Demsetz) pointed out that informational asymmetries form the core of incentive problems in organizations.

The analysis of principal-agent problems with many agents is relatively recent. This strand in the literature takes bilateral agency theory (e.g., Ross, 1973) and extends it to the case of many agents to allow more general analysis of relationships within organizations. This work, represented by Holmstrom (1982) Mookherjee (1984), Green and Stokey (1983), and Nalebuff and Stiglitz (1983), formalizes and extends earlier work on the internal theory of the firm of Coase, Alchian and Demsetz, and Williamson. The principal-agent literature examines explicitly the structure of incentives within an organization which overcome the problem of moral hazard.

It is well-known that it is not possible to design a continuous incentive mechanism which leads to an efficient outcome in an organization in the presence of moral hazard (e.g., Holmstrom, 1982). A number of different approaches have been pursued to attempt to discover efficient solutions to the multiple agent moral hazard problem. Since the negative result on the efficiency of incentive schemes occurs in a single-period, noncooperative game with continuous incentive mechanisms, the approaches to resolving this problem have involved relaxation of one of the conditions of the game.

One approach is to introduce a discontinuous "bonus-penalty" incentive scheme in the single-period noncooperative game. This is Holmstrom's (1982) approach. Holmstrom shows that if a balanced-budget criterion for the organization is relaxed, penalties sufficiently large to induce an efficient outcome are possible. This implies a necessary role for a principal; in particular, separation of management and labor is efficient. Relative performance evaluation is also examined, and is shown to reduce the monitoring costs which must be incurred in the face of uncertainty and risk aversion or limited endowments. This solution occurs under the assumption of deterministic production, risk neutrality, and no problems with agents' endowments.

These caveats about the applicability of an unbalanced budget bonus-penalty incentive scheme are very serious. In reality, one would expect to observe uncertainty in production, risk aversion, and limited endowments on the part of agents. In this situation, relative performance evaluation in the single-period noncooperative game provides an alternative approach to inducing an efficient outcome. Lazear and Rosen (1981), Green and Stokey (1983), and Mookherjee (1984) are papers which examine this approach.

Lazear and Rosen's (1981) paper is the first paper in the relative performance evaluation, or "tournament", literature. Lazear and Rosen compare three compensation schemes; linear piece-rate, discontinuous comparison with a fixed standard, and a rank-order tournament. They show that when the variance of a random component of output common to all agents is large, tournaments lead to more efficient outcomes. This is because the variation in each agent's output is due (mostly) to variation common to all agents; thus relative performance is easily uncovered.

Green and Stokey (1983) show that whether a tournament dominates independent contracts with agents depends on whether a production shock common to all agents exceeds a shock specific to the individual agents. These results indicate that tournaments are not optimal in general, abstracting from the cost of implementing any compensation system. As Green and Stokey suggest, the observed prevalence of tournaments may be due to lesser costs of implementation.

Nalebuff and Stiglitz (1983) report much the same results and consider variations on a number of different assumptions. One important result is that tournaments do not work as well when agents are of different abilities. The incentive for the less able to compete is diminished by the amount of the handicap resulting from lesser ability.

Mookherjee (1984) extends the Grossman and Hart (1983) analysis of the bilateral principal-agent problem to a setting with many agents. The model differs from Green and Stokey (1983) and Nalebuff and Stiglitz (1983) in that a production shock common to all agents is absent. All production shocks are idiosyncratic, i.e., specific to a particular agent. This framework is similar to Holmstrom (1982). Mookherjee shows that independent contracts are optimal when production functions are separable in actions, and agents' idiosyncratic random shocks are independent. Optimality also obtains for very special cases of separable production with non-independent random shocks and nonseparable production with independent random shocks. Rank-order tournaments are optimal when the outputs of different agents convey information about agents' actions only through ordinal rankings. Agents' actions alter the probability of winning, but not the margin of winning, i.e., the correlation between the agents' random shocks is low. Mookherjee shows, however, that relative performance compensation schemes are vulnerable to collusion among the agents. The end result is that tournaments as well as contracts suffer from some weaknesses.

Two other approaches to the design of optimal incentive systems in multiple-agent principal-agent models are to relax either the assumption of the game being noncooperative or static. Aumann (1967) proposed an analysis in which agents enter into binding agreements to enforce a cooperative solution. This is a possible solution to the moral hazard problem in an organization, but solves the problem engendered by noncooperative behavior by effectively removing it. Radner (1981, 1985) has explored repeated principal-agent games, and shown that if discount rates are small, there are equilibria which are approximately efficient.

These propositions do not obtain in the repeated partnership game in which there is no principal. Specifically, if partners in the organization discount the future at all, they may not be able to get close to efficiency. It is possible, however, to show that as partner discount rates move towards zero, they move toward efficient "approximate equilibria". An approximate equilibrium is a combination of strategies for which no one can improve their expected utility by more than a small amount.

All of these approaches to the analysis of efficient compensation systems in organizations assume that all agents and the principal correctly perceive the parameters of their environment. A growing literature in economics and psychology has documented the existence and the effects of misperceptions. Tversky and Kahneman (1974), Slovic et al. (1980), and Einhorn and Hogarth (1981) have investigated and documented the existence of systematic biases in perception. Applications have been made to decision-making under uncertainty (Hey, 1984), and insurance (Kunreuther et al., 1978; Spence, 1977; Polinsky and Rogerson, 1982). The influence of misperceptions has been shown to drastically affect decisions and outcomes. Kleindorfer (1979) and Gaynor (1986) have introduced misperceptions into the analysis of the design of incentives in an organization.

In this paper we generalize and extend our earlier papers to incorporate some of the recent theoretical developments in the economics of organizations. Misperceptions on the part of agents about the production technology and the actions of their peers are introduced into a single-period noncooperative principal-agent game with many agents. It is shown that misperceptions can persist, and that different compensation schemes will be chosen in the presence of misperceptions, and that misperceptions of environmental parameters can significantly affect efficiency and equilibrium outcomes.

We proceed as follows. In the next section, we outline a general framework for the organizational design problem, based on Kleindorfer and Sertel (1979) and related work in organization theory. This framework is then used to pose several problems in incentives and information as these affect organizational behavior and efficiency. Section 3 then considers a few results relating to sharing and wage incentives, under alternative assumptions on the behavior of organizational participants, always assuming complete rationality and perfect information. Section 4 relaxes this perfect information assumption to allow for misperceptions of environmental variables, including parameters in the incentive system and the behaviors of other participants. Section 5 discusses some implications of our results in light of recent empirical work in the theory of the firm and labor economics.

2. ORGANIZATIONAL DESIGN AND INCENTIVES

2.1 Elements of Organizational Design

We model an organization or task-oriented group as a collection of economic agents whose joint contribution to a productive process yields a certain output which each of them considers a good, but whose individual contributions to the group have opportunity costs. The relationship of this model to organizational design is depicted in Figure 1.

The principal, or organization designer, chooses a design $\delta \in \Delta$, e.g. an incentive system, which influences the production decisions (or input choices) $x_i \in X_i$ of group members $i \in N$. Each agent has preferences represented by the utility function $U_i(x,r_i)$ defined on $x \in X = \Pi_N X_i$ and on their remuneration $r_i(x,\delta) \in \mathfrak{R}$ ($\mathfrak{R}$ is the real line), where $x = \{x_i \mid i \in N\} \in X$ is the vector of inputs of agents $i \in N$. Given δ, the agents adjust their decisions, resulting in a collective decision $\underline{x}(\delta)$ and output $F(\underline{x}(\delta),\delta)$.[2] Factor markets provide the basis for agents to determine their opportunity costs, and final product markets determine the price of output, which we will take to be unity here. Agents $i \in N$ are assumed to choose $\underline{x}_i(\delta)$ so as to maximize their utility $U_i(x,r_i)$, $i \in N$. The outcome $\underline{x}(\delta)$ in response to a given design $\delta \in \Delta$ may be thought of as the equilibrium of a strategic-form game among agents N, given δ. Various

[2] F is written here as a function of x and δ for generality. In our later analysis we assume $F = F(x)$ depends only on x.

(cooperative and non-cooperative) solution concepts are of interest in modelling this equilibrium. The principal is assumed to be a risk-neutral agent interested in choosing δ so as to maximize his residual $F(\underline{x}(\delta),\delta) - \Sigma_N r_i(\underline{x}(\delta),\delta)$. We neglect environmental uncertainty in this paper.[3]

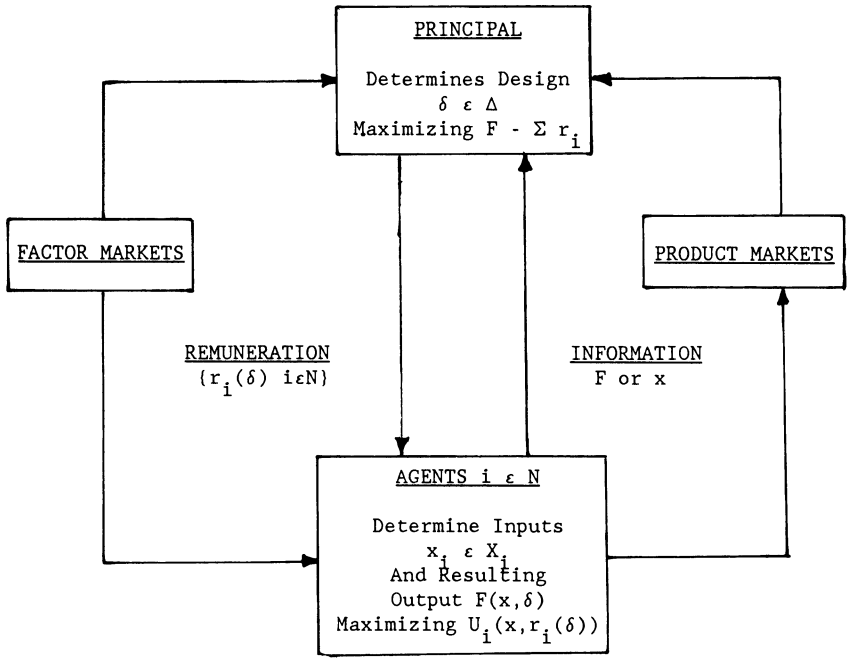


Figure 1: Elements of Organizational Design

From organization theory[4], the design instruments available to the principal are the authority structure (e.g., whether an entrepreneur or the "workers" $i \in N$ themselves choose $\delta \in \Delta$), incentives (the $r_i(x,\delta)$ above), information systems (more about which below), the personnel of the organization (N), and the technology (F). We concentrate here on the design of incentives and compatible

---

[3]Our results are easily extended to the case where the production function and possibly preferences depend on uncertain states of nature--see e.g. the discussion in Kleindorfer and Sertel (1979).

[4]We rely here on the excellent survey of organizational design and behavior contained in Van de Ven and Joyce (1981).

information systems for a fixed authority structure, represented by the principal's interests, and fixed (N,F).

However design instruments are construed, it should be clear that the key to a formal analysis of problems of this type is the appropriate definition of the game-theoretic solution concept specifying the equilibrium outcome $\underline{x}(\delta)$ in response to a particular design δ. In the single-period model of this problem, which is our primary focus, two obvious candidates are the non-cooperative (Nash) and cooperative (Pareto with income transfers among the group N) equilibria. Building on these single-period results, equilibria for the repeated game model can be obtained using the framework and results of Radner (1985). In either case, the key issue here is the definition of the behavioral adjustment process underlying these equilibria. On the one hand, the traditional assumptions of complete and perfect knowledge of all game parameters (U_i, $r_i(\delta)$, x and F) is a natural starting point, and is the basis for our analysis in the next section. On the other hand, it is natural to investigate the impacts of less than perfect knowledge by members $i \varepsilon N$ of these parameters. We will be concerned here only with "misperceptions" of the production function F, whereby agents $i \varepsilon N$ adjust their behaviors in utility-maximizing fashion, while simultaneously estimating the production function F on the basis of observed outcomes. The question of interest is the joint impact of behavioral and perceptual adjustment on efficiency and output.

2.2 The Principal's Design Problem

Summarizing the above, the principal's design problem is the following:

$$\text{(1)} \qquad \underset{\delta \, \varepsilon \, \Delta}{\text{Maximize}} \; F(\underline{x}(\delta),\delta) - \sum_N r_i(\underline{x}(\delta),\delta)$$

subject to:

$$\text{(2)} \qquad U_i(\underline{x}(\delta),\; r_i(\underline{x}(\delta),\delta)) \geq U_{io},\; i \; \varepsilon \; N,$$

where $\underline{x}(\delta)$ is the predicted input vector chosen in response to δ and U_{io} is the reservation utility level for $i \varepsilon N$, as determined by opportunities outside the organization in question. As to the form of $r_i(\delta)$, two polar forms have been of primary interest in the literature, sharing and wage incentives. In the former, $r_i(\underline{x}(\delta),\delta) = r_i(F(\underline{x}(\delta),\delta),\delta)$ depends on $\underline{x}(\delta)$ only through F. Under wage incentives, $r_i(\underline{x}(\delta),\delta) = r_i(\underline{x}_i(\delta),\delta)$ depends only on the input of agent i. Concerning information systems, these must be compatible with the incentive structure chosen. For example, under sharing rules the principal must only observe output F, while under wage incentives he must observe (a one-to-one function of) individual inputs $\{x_i \mid i \in N\}$ in order to pay out the agreed upon remuneration r_i to each agent.[5]

3. DESIGN PROBLEMS UNDER PERFECT INFORMATION

In this section we consider results for the problem (1)-(2) under perfect information and for various solution concepts. To begin with, we make the following regularity assumptions.

Assumptions: The production function $F(x,\delta) = F(x)$ is increasing, continuously differentiable and concave in x, with $F(0) = 0$. Utility functions are of the separable form

$$(3) \qquad U_i(x,r_i) = r_i - V_i(x_i), \quad i \in N,\ x_i \in X_i \subseteq \Re_+,$$

where $\Re_+$ is the non-negative reals, and V_i is an increasing convex, continuously differentiable real-valued function and $V(0) = 0$.

3.1 Noncooperative Sharing Systems; Holstrom's Theorem

Let us begin with a restatement of Holmstrom's (1982) interesting result that sharing incentives which exclude any payments to the principal/designer are inefficient. To this end, let the design set Δ be the set of all sharing incentives:

---

[5]For a more detailed discussion of the compatibility of incentive and information structures, see Kleindorfer and Sertel (1979).

(4) $\Delta := \{S:X\rightarrow\mathbb{R}^N \mid S(x)=\{S_i(x)=S_i(F(x)) \mid i\varepsilon N\},\ \Sigma_N S_i(F)=F,\ F\varepsilon\mathbb{R}_+\}$,

so that each S_i depends only on the output F and the total output value is distributed among the agents N. Associated with each $\delta\varepsilon\Delta$, there is a strategic-form game $\Gamma(\delta) = [N,X,W_i,\delta]$, where, substituting $r_i = S_i$ in (3), the "effective utility functions" W_i induced by $\delta = \{S_i \mid i\varepsilon N\}$ are

(5) $W_i(x,\delta) = S_i(F(x)) - V_i(x_i),\quad i \varepsilon N,\ x \varepsilon X.$

We are interested in comparing the efficiency of various outcomes to the game $\Gamma(\delta)$ with first-best or efficient solutions as determined by:

(6) $$\underset{x\ \varepsilon\ X}{\text{Maximize}}\ [F(x) - \Sigma_N V_i(x_i)].$$

Any solution x* to (6) is Pareto efficient in the sense that any such solution, coupled with appropriate income transfers among agents N, weakly dominates any other $x \varepsilon X$ and feasible transfer payment scheme among N (Kleindorfer and Sertel (1982)). Equation (6) represents the maximum of output value minus opportunity costs of factor inputs x_i.

<u>Theorem 1</u> (Holmstrom (1982)): Let $\delta \varepsilon \Delta$ be arbitrary and let $\underline{x}(\delta)$ be any Nash (noncooperative) equilibrium of $\Gamma(\delta)$, then $\underline{x}(\delta)$ is not Pareto efficient in the sense of (6) when $|N| > 1$. In particular, sharing systems which leave no residual for the principal are not efficient in the sense of (6).

<u>Proof</u>: We merely give Holmstrom's short proof for the case where the S_i are differentiable. In this case, and assuming interior solutions, any Nash solution $\underline{x}(\delta)$ to $\Gamma(\delta)$ must satisfy

(7) $S_i' F_i' - V_i' = 0,\quad i \varepsilon N,$

where $F_i' = \partial F/\partial x_i$, while any Pareto solution to (6) must satisfy

(8) $F_i' - V_i' = 0,\quad i \varepsilon N.$

Since any feasible δ must satisfy $\Sigma_N S_i(F(x)) = F(x)$, $x \varepsilon X$, we also have that

$\Sigma_N S_i' = 1$, for all $x \varepsilon X$. This last requirement is not compatible with (7)-(8) jointly when $|N| > 1$, yielding the desired conclusion. ∇

As noted in the introduction, Holmstrom shows that if budget balancing is relaxed, so that $\Sigma_N S_i(F) \leq F$ is allowed, then sharing incentives can be designed which yield Pareto outcomes as Nash equilibria to $\Gamma(\delta)$. One such Pareto sharing system is of the "forcing function" form described by Harris and Raviv (1978). Namely, $S_i(F) = b_i$ if $F(x) \geq F(x^*)$ and $S_i(F) = 0$ otherwise, where x^* is any solution to (6). Such a sharing system imposes penalties, collected by the principal, for output performance below first-best. As Holmstrom shows, a suitable set of $\{b_i \mid i \varepsilon N\}$ can be determined[6] which will make each $i \varepsilon N$ better off than his best available other alternative and will achieve x^*.

3.2 Nash Bargaining-Cooperative Results

The importance of the principal in avoiding inefficiency in the above problem derives entirely from his design expertise; he contributes no other productive input. The need for the principal as residual owner and "penalty collector" disappears, however, when cooperative behavior on the part of N can be costlessly assured. To formally demonstrate this, we need an appropriate cooperative solution concept for the game $\Gamma(\delta)$. Given the form of the preferences (3), we propose the Nash Bargaining Solution (Nash (1950)), allowing for lump-sum transfers among group members N.[7] In this case, it is easily shown that group members will choose $\underline{x}^*(\delta)$ solving (6). That is, they will choose a first-best solution, assuming that no unproductive principal appropriates any of the usufruct F(x) of production. We summarize this in the following proposition.

---

[6]Namely a set of b_i's such that $\Sigma_N b_i = F(x^*)$ and $b_i > V_i(x_i^*) > 0$, so that each $i \varepsilon N$ is at least as well off taking part in the group's activities as not. Eswaran and Kotwal (1984) show how this scheme can be vulnerable to cheating by the principal. Gaynor (1987) indicates the more limited circumstances under which cheating will occur.

[7]Given the assumption of transferable utility and the form of the preferences (3), other bargaining solutions (see Friedman (1986), Chapter 5) would lead to essentially the same results, with some changes in the ultimate sharing of group surplus. The problem with non-transferable utility remains to be analyzed, however.

Theorem 2: Let δ, $\underline{x}(\delta)$, and $\Gamma(\delta)$ be as in Theorem 1. Consider the Nash Bargaining Solution $\underline{x}^*(\delta)$ to $\Gamma(\delta)$ specified by

$$(9) \qquad \underline{x}^*(\delta) \in \arg\max \{\Pi_N[W_i(x,\delta) + t_i - U_{io}] \mid x \in X_+^*(\underline{t}),\ \underline{t} \in T\},$$

where W_i is given in (5), $T = \{t_i \in \Re \mid \Sigma_N t_i = 0\}$ and $X_+^*(\underline{t})$ is the feasible bargaining set

$$(10) \qquad X_+^*(\underline{t}) = \{x \in X \mid W_i(x,\delta) + t_i \geq U_{io}, i \in N\}.$$

Then (as long as (6) has any solution), $\underline{x}^* = \underline{x}^*(\delta)$ solves (6), i.e. it is first-best, and there exist lump-sum transfers $\underline{t}^* = (t_i^* \mid i \in N)$ such that $(\underline{x}^*, \underline{t}^*)$ Pareto dominates the noncooperative solution $\underline{x}(\delta)$.

Proof: The Nash Bargaining Solution $\underline{x}^*$ characterized by (9) is easily shown to maximize $\Sigma_N W_i(x,\delta)$ over X, which from (4)-(5) implies that $\underline{x}^*$ maximizes (6). Moreover, $\underline{x}^* \in X_+^*(\underline{t})$ and $\underline{t} \in T$, so that $\underline{x}^*$ certainly weakly dominates $\underline{x}(\delta)$ from (9). Finally, $\underline{x}(\delta)$ does not solve (6) by Holmstrom's Theorem 1 above. Thus, some subset of N receives the surplus $\Sigma_N W_i(\underline{x}^*,\delta) - \Sigma_N W_i(\underline{x}(\delta),\delta) > 0$, and strict dominance obtains.[8] ∇

From the above two theorems, we see that an unproductive principal is required only when noncooperative behavior is anticipated. From social and industrial psychology (e.g., Cooper (1975)) we know that cooperation is more likely to obtain when communication and trust are present among agents N, and further enhanced under repeated game situations. More formally, as Radner (1985) has shown[9], when discount rates are sufficiently small (or when gains from cooperation over noncooperation are large), the cooperative solution $\underline{x}^*$ to $\Gamma(\delta)$ can be implemented as a Nash equilibrium to the infinitely repeated game $\Gamma(\delta)\times\ldots\times\Gamma(\delta)\times\ldots$, with payoffs

---

[8] This is similar to the compensation principle in welfare economics. The reader interested in a more detailed analysis of cooperative solutions with lump-sum transfers is referred to Kleindorfer and Sertel (1982), on which this theorem is based.

[9] See also Friedman (1986, Chapter 3) for an introductory survey of other contributions to the repeated game literature.

$$(11) \qquad W_i^{\infty}(x,\delta) = \sum_{k=0}^{\infty} (\alpha_i)^k W_i(x^k,\delta),$$

where $x^k \varepsilon X$ is the collective input chosen by N in period (or sub-game) k and $\alpha_i = 1/(1+d_i) \varepsilon (0,1)$, where d_i is agent i's discount factor. In this sense, the prospect of a "long" association with an organization can be expected to promote cooperative behavior.[10]

3.3 <u>Remarks on Principal-Managed vs. Labor-Managed Enterprises</u>

The above discussion is, of course, not intended as a full commentary on the issue of principal-managed (e.g., entrepreneurial managed) versus labor-managed firms. As indicated in the introduction, and as is apparent from the rich literature on internal organization cited there, this is a much more complicated matter than this sparse formal analysis can capture. In particular, we have neglected entirely the role of capital markets, the potentially productive role of principals in not only designing organizations but also in contributing essential other productive factors such as capital, and the host of issues in monitoring and control which organizational design, by whomever, entails and from which our analysis has abstracted. Even with these caveats, however, it is clear that the issues raised here on cooperative vs. noncooperative internal adjustment by group/organizational agents are central to any analysis of organizational and incentival design. Moreover, the centrality of gains to cooperation and their relationship to cut-off discount factors by agents has been a central point in the literature on labor-managed vs. capitalist-managed firms. Indeed, one of the basic arguments of this literature has been that workers tend to have higher discount factors than capitalists and this induces them to maintain inefficiently low levels of capitalization (e.g., through retained earnings) in the firm in a dynamic environment. This argument is then used in various ways to explain the

[10]This very point is argued to be the crux of the Japanese productivity miracle, based on long-term employment and cooperation-communication inducing incentives and organizational processes. See Tomer (1985).

predominance of labor-managed firms only in certain areas (such as professional partnerships), where the gains to cooperation are very high and monitoring costs are also high.[11] In other areas, e.g. manufacturing and capital-intensive sectors, successful labor-managed firms are a rarity, perhaps partly because worker discount rates are high and their expected tenure in the firm is short and perhaps partly because gains to cooperation are low. In either case, the above analysis would suggest a tendency toward noncooperative behavior, with a resulting requirement that the organization be designed by a principal/residual owner.

4. MISPERCEPTIONS AND GROUP INCENTIVES

4.1 Consistent Non-Cooperative Equilibrium

Most previous work on incentives has assumed that agents possess accurate and common estimates of production technology and other environmental parameters. Thus, in defining the noncooperative equilibrium to the game $\Gamma(\delta)$ in the previous section, it is assumed that the effective utility functions are given by (5), with F(x) (and, of course, also S_i) accurately understood by all agents and by the principal. In this section, we wish to investigate the consequences of agents' misperceptions of F(x) on equilibrium outcomes. Our purpose is to demonstrate that a noncooperative equilibrium still exists under fairly general conditions, even when agents misestimate the production function F(x). Similar results can be shown for cooperative solutions and for other environmental parameters of interest (e.g., misestimates of the opportunity cost $V_i(x_i)$ of the inputs x_i), but we will concentrate here only on noncooperative adjustment and misperceptions of F.

The motivation for this problem should be clear. There is a very rich empirical literature, beginning with Simon (1957) and Cyert and March (1963), which supports the bounded rationality hypothesis, that agents are intendedly

[11] For a fuller discussion of comparative results on hierarchical vs. labor-managed firms, see Ireland and Law (1982), Sertel (1982) and Cable and Fitzroy (1980).

rational but have limited information processing abilities. These limitations lead to processing heuristics (such as satisficing) and systematic biases. This may be particularly expected in complex environments such as those envisioned here, where agents need to precisely understand the impact of their actions on their payoffs as determined by incentives. When these incentives depend, as they do under generalized output sharing, on the production technology (i.e., when payments are of the form $S_i(F)$), this implies that agents accurately assess the impact of at least their own inputs on output F. That this is a strong requirement can easily be visualized by considering its implications for employees of a large firm, in which employees are given profit-sharing incentives. The problem (which arises in maximizing (5)) facing an individual employee in predicting the impact of changes in his/her own input on the total output F(x) of the firm staggers the imagination. In such a case, it seems sensible that employees would form crude, perhaps inaccurate, estimates of their relationship to aggregate output. What we study here is whether inaccuracies in their perceptions of F can be sustained in equilibrium.

We restrict attention again to sharing incentives of the form (3) but replace the effective utility functions (5) with

(12) $$W_i(x,y,\delta) = S_i(G_i(x,y)) - V_i(x_i),\ i \in N,\ x \in X,\ y \in X,$$

where $G_i: X \times X \to \Re$ is agent i's <u>perceived production function</u>, $i \in N$. The interpretation of $G_i(x,y)$ is the following. The input vector x is as before the collective choice variable of interest. The input vector y is a reference point for x, possibly the collective input vector in the previous period. We assume that when x = y perceived and actual output are the same, i.e. G_i satisfies

(13) $$G_i(x,x) = F(x),\quad \forall\ x \in X,\ i \in N.$$

An intuitive interpretation of (13) is that each agent forms a Taylor series approximation of F around y, so that

(14) $$G_i(x,y) = F(y) + \sum_{k=1}^{K} \Phi_i^k(x - y)$$

here $\Phi_i^k(0) = 0$. Φ_i^k can be interpreted as agent i's estimate (perhaps erroneous) of the vector of kth order partials of F with respect to x and K is the order of agent i's approximation. F(y) is output realized at the reference input level y.

We wish to determine when the game with utility functions W_i in (12) has a noncooperative equilibrium. However, not just any equilibrium will do. To qualify as a sustainable equilibrium, it seems reasonable to require that the payout $S_i(G_i(x,y))$ expected by agents (whatever their perceived production functions are) be at least equal to their actual payout. Interpreting the reference point y as agents' previous behavior, we state this condition as follows:

Definition: Consider the normal form game $\Gamma(\delta) = [N,X,Y,W_i,\delta]$, with $X = Y$ and $W_i(x,y,\delta)$ given by (12). A consistent (noncooperative) equilibrium $(\underline{x}(\delta),\underline{y}(\delta))$ to $\Gamma(\delta)$ is a collective behavior $\underline{x}$ and reference behavior $\underline{y}$ such that

(15) $\underline{x} = \underline{y} = \underline{x}(\delta)$

(16) $G_i(\underline{x},\underline{y}) = F(\underline{x}),\ i \in N.$

(17) $\underline{x}_i \in \arg\max_{x_i \in X_i} W_i(x_i,\underline{x}_{-i},\underline{y},\delta),\ i \in N,$

where $\underline{x}_{-i} = \{\underline{x}_j \mid j \in N\backslash i\}$ is the i-exclusive behavior at equilibrium. Condition (17) is just the Nash condition; (16) is the indicated consistency condition, i.e., (16) implies $S_i(G_i(x,y)) = S_i(F(x))$; (15) is a further equilibrium condition that the reference point for the approximating functions G_i be the same as the actual behavior at equilibrium.

4.2 Sufficient Conditions for the Existence of Consistent Equilibria

We now show that a wide class of incentive designs $\delta \in \Delta$ and perceived production functions give rise to games $\Gamma(\delta)$ with consistent equilibria.

Theorem 3: Fix $\delta \in \Delta$. Assume that G_i satisfies (13), $i \in N$, that S_i is continuous, concave and increasing, that G_i is concave in x_i for any fixed x_{-i}

and y. Assume further that X is nonempty, compact and convex.[12] Then the game $\Gamma(\delta)$ has a consistent equilibrium.

<u>Proof</u>: We first enlarge the game $\Gamma(\delta)$ as follows. We assume that there are utility maximizing agents $i \in M = \{n+1, \ldots, 2n\}$, with the (n+i)th such agent determining y_i in noncooperative fashion so as to maximize

$$(18) \qquad W_i(x,y,\delta) = -(x_i - y_i)^2, \; i \in M.^{13}$$

Now consider the game with $N \cup M$ players, with utility functions W_i given by (12) and (18) and behavior spaces X_i for $i \in N$ and Y_i for $i \in M$. It is clear that W_i is concave for $i \in M$ from (18). For $i \in N$, the fact that S_i is concave increasing and G_i is concave in x_i implies that $S_i(G_i(x,y))$ is concave in x_i for any fixed (x_{-i}, y). Thus, given the convexity of V_i, W_i is concave in x_i. Applying Kakutani's Fixed Point Theorem yields the assertion in usual fashion (since obviously the best-response mapping for $i \in M$ is always $y_i \quad x_i$ by (18)). ∇

4.3 <u>Illustrative Example on the Effect of Misperceptions on Efficiency</u>

An example may serve to illustrate the nature of the above theorem. Consider pure sharecropping incentives of the form $S_i(F) = \lambda_i F$, where $0 \le \lambda_i \le 1$, is agent i's sharing constant (with $\Sigma_N \lambda_i = 1$). Let agent i's perceived production function be a first-order Taylor series approximation of the form (14):

$$(19) \qquad i(x,y) = F(y) + b_i(x_i - y_i) + \Psi_i(x_{-i} - y_{-i}),$$

where $b_i \in \Re$ is agent i's estimate of his marginal productivity and Ψ_i is his

---

[12] Other conditions on G_i and X would suffice. The basic requirement is that the maximization problem embodied in (17) always gives rise to a maximizing x_i in a compact, convex subset of X_i.

[13] (18) is an arbitrarily chosen function ensuring that the estimation process leads to $\underline{x} = \underline{y}$. A more intuitive interpretation would be one of dynamic revision of estimates, e.g.,

$$G_i(x^{t+1}, x^t, \ldots, x^o) = F(x^t) + \sum_{\tau=0}^{t+1} \sum_{k=1}^{k} \phi_i^{k\tau}(x^{\tau+1} - x^\tau),$$

where the t's and τ's are time indices.

estimate of the aggregate marginal impact of other agents' changes in inputs. Assuming strict convexity of V_i and an interior solution, these assumptions lead to the following first-order conditions for (18):

(20) $V_i'(x_i) = \lambda_i b_i, \; i \in N,$

so that the unique consistent equilibrium for $\Gamma(\underline{\lambda})$ is

(21) $\underline{x}_i(\underline{\lambda}) = V_i'^{-1}(\lambda_i b_i), \; i \in N.$

It is interesting to compare these consistent equilibria with the perfect information noncooperative equilibria determined by (7), which in this case (where $S_i' = \lambda_i$) would be

(22) $\underline{x}_i(\underline{\lambda}) = V_i'^{-1}(\lambda_i F_i'), \; i \in N.$

When $b_i = F_i'$, $i \in N$, i.e., no misperceptions, these would be identical, but not in general. In particular, comparing (21)-(22) and noting that $V_i'^{-1}$ is increasing by convexity, we see that when every agent overestimates (resp., underestimates) his marginal product (i.e., when $b_i \geq F_i'$ or $b_i \leq F_i'$, $i \in N$), the consistent equilibrium under misperceptions will yield higher (resp., lower) inputs and utilities (and, of course, output and incentive payments) to each agent than under perfect information. Under such misperceptions, agents work harder (resp., less) and their expectations of increased (resp., decreased) marginal benefits are fulfilled in equilibrium.[14] This is possible since the consistency conditions restrict choice to $\underline{x} = \underline{y}$, which implies $G_i(x,x) = F(x)$.

Several generalizations of the above are straightforward. First, consistent equilibria are easily shown to exist if perceived production functions are "estimated" not just from a single reference point (e.g., last period's behavior), but from a finite set of reference points (e.g., the last h periods'

[14]For a series of other, more detailed, examples in a related framework of the effect of misperceptions on output and utilities at consistent equilibria, see Kleindorfer (1979). For a proof of these results without the consistency conditions, see Gaynor (1986).

behaviors). We need only define G_i as $G_i = G_i(x,y^1,\ldots,y^h)$, where $y^1,\ldots,y^h$ are the reference points in question, and where (13) becomes $G_i(x,x,\ldots,x) = F(x)$. Then introduce, as in (19), a set of h×n fictive players each maximizing over some y_i^j, $i \varepsilon N$, $j \varepsilon \{1,\ldots,h\}$, with utility functions

$$(23) \qquad W_i(x,y,\delta) = -(y_i^{j-1} - y_i^j)^2.$$

The game with utility functions (23) and (13) will then yield as before a consistent equilibrium $\underline{x}(\delta) = \underline{x} = \underline{y}^1 = \ldots = \underline{y}^h$, with $G_i(\underline{x},\underline{y}^1,\ldots,\underline{y}^h) = F(\underline{x})$.

The efficiency implications of these results are interesting. Referring to (21) and (22), it is clear that when every agent overestimates his marginal product ($b_i > F_i'$), welfare is higher at the consistent equilibrium with misperceptions than at the perfect information equilibrium. The converse is true for underestimation ($b_i < F_i'$). This implies that not only do optimists think that the world is a better place, they make it a better place. It also has some implications for the sort of messages a designer may wish to include in δ. We have not explored the welfare implications of mixed (both over- and under-estimation) misperceptions on the part of agents. If the agents are identical in every respect except for their estimations of the technology G_i, it is clear that whether welfare is higher or lower with misperceptions than with perfect information will depend on the number of agents who over- and under-estimate, and the strength of their misperceptions. Referring back to the principal's design problem, his welfare is defined by

$$(24) \qquad F(\underline{x}(\delta)) - \Sigma_N S_i(F(\underline{x}(\delta))),$$

from (1). If (24) is increasing (at least locally) in x, then the principal is clearly better off when the agents are better off. If this is not the case, a potential for conflict exists. If the organization's revenues do not rise more rapidly than agents' payoffs, this conflict will exist. As an example, if demand is slack for a firm, F(x) may not be an increasing function of x. In this case management must reduce payouts to the agents, either through the functions determining the r_i, or through messages which influence agents' perceptions.

This investigation awaits further research.

Other topics of some interest would include the stability of the consistent equilibria determined above, especially given the intuitive interpretation of these equilibria as being derived from perceived production functions identified on the basis of past observed outputs and behaviors. More generally, this dynamic interpretation deserves, of course, a repeated game treatment, whereby agents are aware in selecting their behaviors of the consequences of their decisions for not only this period's payoff but also, through their fixed production function identification process embodied in G_i, on their future payoffs. In the same vein, misperceptions could be modeled here as uncertainty about a parameter, say α, in the production function $F(x,\alpha)$, with each agent i having a subjective probability distribution on α, which is identified over time via appropriate (e.g., Bayesian) methods. This could include misperceptions about other agents' actions $\underline{x}_i$ or the design δ as well as the technology. These extensions will have to await future research. However, the approach taken here, and especially the consistency conditions imposed, may indicate a fruitful way of pursuing these extensions.

5. CONCLUDING REMARKS AND REVIEW OF RELEVANT EMPIRICAL STUDIES

This paper has reviewed organizational and group incentive problems using a multiple-agent principal-agent framework. Our main objective has been to explore the effect of introducing misperceptions by agents of environmental parameters on the existence and efficiency of equilibrium outcomes. Inter alia, we have shown for a single-period model and rather simple perception-adjustment processes that misperceptions can persist at equilibrium, with non-trivial consequences for efficiency and welfare of both the principal/designer and agents. These results are interesting since economic theory has consistently rejected the possibility of persistent mistakes by rational agents, yet empirical research in economics and other disciplines has increasingly unearthed evidence in seeming support of persistent biases. Our research indicates the possibility of rationalizing such empirical results within the context of received economic theory.

We cite for interest three recent empirical studies which may be interpreted as consistent with the results presented above. Nerlove (1983), in investigating expectations and realizations for a sample of French and German firms, found that all firms reported expectations in the no-change category much more frequently than realizations. In addition, the German firms consistently overestimated the balance between price increases and decreases and consistently underestimated the balance between demand increases and decreases. These findings are consistent with a hypothesis of sustained misperceptions (i.e. biased estimates) of the demand process.

The literature in labor economics concerned with explaining the causes of strikes has also produced some relevant evidence. Fudenburg, Levine and Ruud (1984), in investigating the causes of strikes due to asymmetric information (as theorized by Ashenfelter and Johnson (1969) and Hayes (1984)) report that strikes are more likely to occur when a firm is doing poorly relative to the economy. Gunderson, Kervin and Reid (1986), in investigating the determinants of strike incidence in Canada, find that strikes were more likely in firms which had experienced high growth in employment in the previous period. These results are consistent with our specification that conflict will result if the principal's objective function is non-increasing in **x**.

Concerning future research, it would be interesting to extend these results to the case of uncertainty and Bayesian perceptual adjustment, given observed history. Also, the indicated extensions of our results in the Radner repeated game context, allowing for uncertainty and Bayesian estimates of uncertain parameters, should be rather interesting. Finally, applications of these ideas to specific problems in internal organization, labor economics and the positive theories of firm behavior in reacting to environmental shocks could be fruitful in explaining the macro consequences of boundedly rational "human nature as we know it."

REFERENCES

Alchian, A. and Demsetz, H. (1972) "Production, Information Costs, and Economic Organization," American Economic Review 62:777-95.

Arrow, K. (1951) Social Choice and Individual Values New York: Wiley.

Ashenfelter, O. and Johnson, G. E. (1969) "Bargaining Theory, Trade Unions, and Industrial Strike Activity," American Economic Review 59:35-49.

Aumann, R. (1967) "A Survey of Cooperative Games without Side Payments," in Shubik, M. (ed.) Essays in Mathematical Economics in Honor of Oskar Morgenstern, Princeton: Princeton University Press.

Berle, A. and Means, G. (1932) The Modern Corporation and Private Property, New York: Macmillan.

Cable, J. R. and Fitzroy, F. (1980), "Productive Efficiency, Incentives and Employee Participation: Some Preliminary Results for West Germany", Kyklos, 33: 100-121.

Coase, R. (1937) "The Nature of the Firm," Economica 4:386-405.

Cooper, C. L. (1975) Theories of Group Processes, New York: Wiley.

Cyert, R. and March, J. (1963) Behavorial Theory of the Firm, Englewood Cliffs: Prentice-Hall.

Einhorn, H. and Hogarth, R (1981) "Behavioral Decision Theory: Processes of Judgement and Choice," Review of Psychology 32:53-88.

Eswaran, M. and Kotwal, A. (1984) "The Moral Hazard of Budget-Breaking," Rand Journal of Economics 15:578-81.

Friedman, J. W. (1986) Game Theory with Applications to Economics, New York: Oxford University Press.

Fudenburg, D. D., Levine, D. K. and Ruud, P. A. (1984) "Strike Activity and Wage Settlements," Working Paper University of California, Los Angeles, CA.

Gaynor, M. (1986) "Misperceptions, Moral Hazard and Incentives in Groups," Managerial and Decision Economics, 7:279-82.

Gaynor, M. (1987) "The Presence of Moral Hazard in Budget-Breaking," unpublished paper.

Gordon, R. A. (1945) Business Leadership in the Large Corporation, Washington, D.C., reprinted by the University of California Press, Berkeley, 1961.

Green, J. and Stokey, N. (1983) "A Comparison of Tournaments and Contracts," Journal of Political Economy 91: 349-364.

Grossman, S. and Hart, O. (1983) "An Analysis of the Principal-Agent Problem, Econometrica 51:7-46.

Groves, T. (1973) "Incentives in Teams", Econometrica 41:617-631.

Groves, T. and Ledyard, J. (1977) "Optimal Allocation of Public Goods: A Solution to the 'Free-Rider' Problem," Econometrica 45: 783-809.

Gunderson, M., Kervin, J. and Reid, F. (1986) "Logit Estimates of Strike Incidence from Canadian Contract Date," Journal of Labor Economics 4:257-76.

Harris, M. and Raviv, A. (1978), "Some Results on Incentive Contracts", The American Economic Review, 68:1, 20-30.

Hayes, B. (1984) "Unions and Strikes with Asymmetric Information," Journal of Labor Economics 2:57-83.

Hey, J. (1984) "The Economics of Optimism and Pessimism," Kyklos 37:181-205.

Holmstrom, B. (1982) "Moral Hazard in Teams," Bell Journal of Economics 13:324-40.

Ireland, N. J. and Law, P. J. (1982) The Economics of Labor-Managed Enterprises, London: Croom Helm.

Kahneman, D. and A. Tversky (1979) "Prospect Theory, An Analysis of Decision Under Risk," Econometrica 47:263-291.

Kleindorfer, P. R. (1979) "Informational Aspects of Enterprise Design Through Incentives," in Klaus Krippendorff (ed.) Communication and Control in Society, New York: Gordon and Breach.

Kleindorfer, P. R. and Sertel, M. R. (1979) "Profit-Maximizing Design of Enterprises through Incentives," Journal of Economic Theory 20:318-39.

Kleindorfer, P. R. and Sertel. M. R. (1982) "Characterizing Cooperative Solutions to Enterprise Design Problems", 227-32 in Sertel (1982).

Kunreuther, H., et. al. (1978) Disaster Insurance Protection: Public Policy Lessons, New York: Wiley.

Lazear, E. and Rosen, S. (1981) "Rank Order Tournaments as Optimum Labor Contracts," Journal of Political Economy 89: 841-864.

Leibenstein, H. (1966) "Allocative Efficiency vs. X-Efficiency," American Economic Review 56:392-415.

Marris, R. (1964) The Economic Theory of "Managerial" Capitalism, New York: Macmillan.

Marschak, J. and Radner, R. (1972) Economic Theory of Teams, New Haven: Yale University Press.

Mookherjee, D. (1984) "Optimal Incentive Schemes with Many Agents," Review of Economic Studies 51: 433-446.

Nalebuff, B. and Stiglitz, J. (1983) "Prizes and Incentives: Towards a General Theory of Compensations and Competition," Bell Journal of Economics 14:21-43.

Nash, J. F. Jr. (1950) "The Bargaining Problem," Econometrica 18:155-62.

Nerlove, M. (1983) "Expectations, Plans, and Realizations in Theory and Practice," Econometrica 51:1251-1279.

Polinsky, A. and Rogerson, W. (1982) "Products Liability, and Consumer Misperceptions, and Market Power," National Bureau of Economic Research, Working Paper No. 937.

Radner, R. (1981) "Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship," Econometrica 49:1127-48.

Radner, R. (1985) "Repeated Principal - Agent Games with Discounting," Econometrica 53:1173-98.

Ross, S. (1973) "The Economic Theory of Agency: The Principal's Problem," American Economic Review 63:134-9.

Sertel, M. R. (1982), Workers and Incentives, Amsterdam: North Holland.

Simon, H. A. (1957) Models of Man, New York: Wiley.

Slovic, P., Fischhoff, B. and Lichtenstein, S. (1980) "Facts and Fears: Understanding Perceived Risk," in Schwing, R. and Albers, W. (eds.) Societal Risk Assessment, New York: Plenum Press.

Spence, A. (1977) "Consumer Misperceptions, Product Failure, and Producer Liability," Review of Economic Studies 44:561-72.

Tomer, J. F. (1985) "Working Smarter the Japanese Way: The X-Efficiency of Theory Z Management," 199-224 in P. R. Kleindorfer (ed.) The Management of Productivity and Technology in Manufacturing, New York: Plenum Press.

Tversky, A. and Kahneman, D. (1974) "Judgment under Uncertainty: Heuristics and Biases," Science 185:1124-31.

Van de Ven, A. H. and Joyce, W. F. (1981) Perspectives on Organization Design and Behavior, New York: Wiley.

Williamson, O. E. (1963) "Managerial Discretion and Business Behavior," American Economic Review 53:1032-57.

Williamson, O. E. (1975) Market and Hierarchies: Analysis and Antitrust Implications, New York: Free Press.

Williamson, O. E. (1986) The Institutions of Modern Capitalism, New York: Free Press.

Intertemporal Sharecropping: A Differential Game Approach

G. Feichtinger and G. Sorger

Summary: The sharecropping paradigm describes a special principal-agent situation. Most models in this field are not formulated in a dynamic framework. Although multistage game models have been considered in this context, a treatment of the intertemporal problem in continuous time is still missing. The present paper can be considered as a first step into this direction. We consider a dynamic, non-stochasti sharecropping model to analyze the strategic competition of a principal and an agent. Depending on the information structure and the kind of contract between the landlord and the farmer different solution concepts for the dynamic game are considered. Especially we shall discuss several noncooperative and cooperative solutions. Because of the inherent asymmetry of the situation, the Stackelberg equilibrium is of special importance for the proposed approach to sharecropping.

Table of Contents:

1. Differential Games

1.1 Problem Formulation and Open Loop Information Structure

Differential game theory provides a framework for the analysis of the interaction of economic agents. It is the appropriate mathematical tool to determine optimal decisions under various behavioural assumptions.

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

In this section we briefly present a dynamic games set up. We discuss the open-loop information structure as well as some solution concepts (see, e.g., Basar and Olsder, 1982).

Let $\underline{x}(t) \in \mathbb{R}^n$ be the state of a system at time t which evolves on the time interval [0,T] according to the ordinary differential equation

$$\dot{\underline{x}}(t) = \underline{f}(\underline{x}(t), \underline{u}_1(t), \ldots, \underline{u}_N(t), t), \ \underline{x}(0) = \underline{x}_0,$$

where $\underline{u}_i(t) \in \Omega_i \subseteqq \mathbb{R}^{m_i}$ is the control for player i = 1, ..., N. The payoff to player i is given by

$$J_i(\underline{u}_1 \ldots, \underline{u}_N) = \int_0^T e^{-rt} F_i(\underline{x}, \underline{u}_1, \ldots, \underline{u}_N, t) dt,$$

where each player has the same discount rate r. For simplicity of notation we omit the time arguments.

The concept of information structure plays an important role. As to the choice of the control variables several possibilities are available, depending on the information structure assumed for the game. Since we shall consider only open-loop strategies we restrict ourselves to the explanation of this concept. An open-loop strategy $\underline{u}_i = \underline{u}_i(t, \underline{x}_0)$ depends only on time and initial state. Thus, in a game with open-loop information the players commit themselves to time paths before the start of the game.

1.2 Solution Concepts

In game theory there is no unique solution concept. From now on we consider only *two-person* games. The *Nash solution* is secure in the sense that no player can obtain a better payoff by unilaterally deviating from his Nash strategy as long as the other player sticks to his Nash strategy:

Definition 1. A Nash equilibrium $(\underline{u}_1^N, \underline{u}_2^N)$ is a pair of strategies such that

$$J_1(\underline{u}_1, \underline{u}_2^N) \leq J_1(\underline{u}_1^N, \underline{u}_2^N) \quad \textit{for all admissible} \quad \underline{u}_1$$

$$J_2(\underline{u}_1^N, \underline{u}_2) \leq J_2(\underline{u}_1^N, \underline{u}_2^N) \quad \textit{for all admissible} \quad \underline{u}_2.$$

Another interesting solution concept is the *Stackelberg equilibrium* which is characterized by asymmetric information. The first player, the follower, maximizes his objective for all possible values of $\underline{u}_2$. This defines a reaction function $R_1(\underline{u}_2)$. Knowing this reaction function the second player, i.e. the leader, optimizes his criterion.

Definition 2. A Stackelberg equilibrium $(\underline{u}_1^S,\underline{u}_2^S)$ *with player 1 as follower and player 2 as leader is a pair of strategies such that*

$$J_2(R_1(\underline{u}_2),\underline{u}_2) \leqq J_2(R_1(\underline{u}_2^S),\underline{u}_2^S)$$

for all admissible $\underline{u}_2$, *and* $\underline{u}_1^S = R_1(\underline{u}_2^S)$. *Moreover the reaction function* $R_1(u_2)$ *is implicitly defined by*

$$J_1(R_1(\underline{u}_2),\underline{u}_2) = \max_{\underline{u}_1} J_1(\underline{u}_1,\underline{u}_2)$$

Whereas the above solution concepts are noncooperative, the following *Pareto equilibrium* is a cooperative solution.

Definition 3. A Pareto equilibrium $(\underline{u}_1^P,\underline{u}_2^P)$ *is a pair of strategies such that either for all admissible* $\underline{u}_1$, $\underline{u}_2$ *it holds that*

$$J_i(\underline{u}_1,\underline{u}_2) = J_i(\underline{u}_1^P,\underline{u}_2^P) \quad \text{for} \quad i = 1,2$$

or there exists $j\varepsilon\{1,2\}$ *with*

$$J_j(\underline{u}_1,\underline{u}_2) < J_j(\underline{u}_1^P,\underline{u}_2^P).$$

A Pareto solution has the property that no player can improve his payoff without diminishing the return of his opponent. It is well-known that each Pareto solution can be obtained by maximizing a weighted sum of the objective functionals of both players:

$$J(\underline{u}_1,\underline{u}_2) = J_1(\underline{u}_1,\underline{u}_2) + \mu J_2(\underline{u}_1,\underline{u}_2)$$

with $\mu\varepsilon(0,\infty)$.

What value for μ is chosen depends on the following bargaining process. In the first step both players announce *threat strategies* $\underline{w}_1$, $\underline{w}_2$. In the second step the players agree on jointly maximizing the product

$$[J_1(\underline{u}_1,\underline{u}_2) - J_1(\underline{w}_1,\underline{w}_2)][J_2(\underline{u}_1,\underline{u}_2) - J_2(\underline{w}_1,\underline{w}_2)]$$

such that $J_i(\underline{u}_1,\underline{u}_2) \geqq J_i(\underline{w}_1,\underline{w}_2)$ for $i = 1,2$. The result is a pair of strategies $(\underline{v}_1,\underline{v}_2)$ which depends on the threat strategies $(\underline{w}_1,\underline{w}_2)$. Therefore we can write

$$J_i(\underline{v}_1,\underline{v}_2) = I_i(\underline{w}_1,\underline{w}_2) \quad \text{for} \quad i = 1,2.$$

A last step consists of the determination of optimal threat strategies such that

$$I_1(\underline{u}_1^*,\underline{u}_2^*) \geqq I_1(\underline{w}_1,\underline{u}_2^*), \quad I_2(\underline{u}_1^*,\underline{u}_2^*) \geqq I_2(\underline{u}_1^*,\underline{w}_2)$$

for all admissible pairs of threats (w_1,w_2). $(\underline{u}_1^*,\underline{u}_2^*)$ is called a pair of optimal threat strategies, and the corresponding Pareto solution $\underline{v}_i(\underline{u}_1^*,\underline{u}_2^*)$ (i = 1,2) is the Nash bargaining equilibrium (see Liu, 1973, Pohjola, 1984).

2. The Model

A landlord owns land which is divided in cultivated area and wilderness. He offers a farmer a portion of the agricultural acreage for tillage and utilizes the rest for himself. In return for the rights of usufruct the farmer cultivates new acreage. Moreover, the landlord may also make arable the soil by himself. Cultivated land changes to desert with a constant rate. In the following this situation is described as a differential game.

Denote by $A(t)$ the agricultural acreage at time t and by δ the natural decay rate of land. Let $u_1(t)$ be the rate of cultivation of the farmer (player 1) and $u_2(t)$ be the cultivation rate of the landlord (player 2) at time t. Then the system dynamics is given by

$$\dot{A}(t) = u_1(t) + u_2(t) - \delta A(t). \tag{1}$$

Player 2 allocates at each time the cultivated area between the famer and himself. The corresponding portions are denoted by $1 - v_2(t)$ and $v_2(t)$, respectively. It is reasonable to assume that the yield is proportional to the acreage. The cultivation causes costs which for simplicity are supposed to be quadratic. Moreover, a constant duration of the game, T, is assumed. With r being a nonnegative discount rate the objective functionals for the two players can be written as:

$$J_1 = \int_0^T [(1-v_2(t))A(t) - \frac{c_1}{2}u_1(t)^2]e^{-rt}dt \tag{2a}$$

$$J_2 = \int_0^T [v_2(t)A(t) - \frac{c_2}{2}u_2(t)^2]e^{-rt}dt. \tag{2b}$$

The problem is summarized as follows:

$$\max_{u_1} J_1, \quad \max_{u_2,v_2} J_2 \qquad (3a) \qquad\qquad \dot{A} = u_1 + u_2 - \delta A,\ A(0) = A_0 \qquad (3b)$$

$$u_1 \geq 0,\ u_2 \geq 0 \qquad (3c) \qquad\qquad 0 \leq v_2 \leq 1. \qquad (3d)$$

Note that here and in the following the time arguments are suppressed.

3. Nash Equilibrium

To calculate the noncooperative Nash solution an optimal control problem is solved for each player.

3.1 Problem of Player 1

With the current-value Hamiltonian

$$H_1 = (1-v_2)A - \frac{c_1}{2}u_1^2 + \lambda_1(u_1+u_2-\delta A)$$

the necessary optimality conditions are

$$u_1 = \arg\max_{u_1\geq 0} H_1 \Rightarrow u_1 = \begin{cases} \lambda_1/c_1 & \text{for } \lambda_1 \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$\dot{\lambda}_1 = r\lambda_1 - \partial H_1/\partial A = (r+\delta)\lambda_1 - (1-v_2) \tag{5}$$

$$\lambda_1(T) = 0. \tag{6}$$

From (5) and (6) follows that the adjoint variable λ_1 of player 1 is given by

$$\lambda_1(t) = \int_t^T e^{-(r+\delta)(\tau-t)}(1-v_2)d\tau. \tag{7}$$

Therefore, λ_1 is always nonnegative, and (4) yields the reaction function

$$u_1(t) = \frac{1}{c_1}\int_t^T e^{-(r+\delta)(\tau-t)}(1-v_2)d\tau. \tag{8}$$

3.2 Problem of Player 2

The Hamiltonian of player 2 is given by

$$H_2 = v_2A - \frac{c_2}{2}u_2^2 + \lambda_2(u_1+u_2-\delta A).$$

The necessary optimality conditions are

$$u_2 = \arg\max_{u_2\geq 0} H_2 \Rightarrow u_2 = \begin{cases} \lambda_2/c_2 & \text{for } \lambda_2 \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$v_2 = \arg\max_{0\leq v_2\leq 1} H_2 \Rightarrow v_2 = \begin{cases} 1 \\ \text{undefined} \\ 0 \end{cases} \quad \text{for } A \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} 0 \tag{10}$$

$$\dot{\lambda}_2 = r\lambda_2 - \partial H_2/\partial A = (r+\delta)\lambda_2 - v_2 \tag{11}$$

$$\lambda_2(T) = 0. \tag{12}$$

The solution of (11) and (12) is

$$\lambda_2(t) = \int_t^T e^{-(r+\delta)(\tau-t)} v_2 d\tau \geq 0. \tag{13}$$

From this and (9) follows

$$u_2(t) = \frac{1}{c_2} \int_t^T e^{-(r+\delta)(\tau-t)} v_2 d\tau. \tag{14}$$

Obviously for $A_0 > 0$ we have that $A(t) > 0$ for all $t > 0$. To show this property also for $A_0 = 0$ we use

$$\dot{A}(0) = u_1(0) + u_2(0) = \int_0^T e^{-(r+\delta)t} \left(\frac{1-v_2}{c_1} + \frac{v_2}{c_2} \right) dt$$

$$\geq \min\left(\frac{1}{c_1}, \frac{1}{c_2}\right) \frac{1}{r+\delta}(1-e^{-(r+\delta)T}).$$

From $A(t) > 0$ we conclude by (10) that

$$v_2(t) = 1 \quad \text{for all} \quad t > 0. \tag{15}$$

Substitution of (15) in (8) and (14) yields

$$u_1(t) = 0 \quad \text{for all} \quad t > 0 \tag{16}$$

$$u_2(t) = \frac{1}{c_2(r+\delta)}(1-e^{-(r+\delta)(T-t)}). \tag{17}$$

Thus we have proved the following result.

Proposition 1. The noncooperative Nash equilibrium of the differential game (3) is given by (16), (17) and (15).

It is obvious that the farmer's Nash profit is zero, whereas the landlord earns always a positive amount: $0 = J_1 < J_2$. Clearly the Nash solution is of no practical importance. However, it has been calculated to compare it with the Stackelberg equilibrium which is dealt with in the following section.

4. Stackelberg Solution

We now turn to the Stackelberg case with the landlord as leader and the farmer as follower.

4.1 Necessary Optimality Conditions

Taking into consideration the reaction (4) - (6) of the farmer the landlord faces the following two-state variable optimal control problem

$$\max_{u_2,v_2} \int_0^T (v_2A - \frac{c_2}{2}u_2^2)e^{-rt}dt \tag{18a}$$

$$\dot{A} = \frac{\lambda_1}{c_1} + u_2 - \delta A, \; A(0) = A_0 \tag{18b}$$

$$\dot{\lambda}_1 = (r+\delta)\lambda_1 - (1-v_2), \; \lambda_1(T) = 0. \tag{18c}$$

Note that the costate of player 1, λ_1, acts as an additional state variable of player 2 (see, e.g., Feichtinger and Hartl, 1986).

The Hamiltonian of problem (18) is given as

$$H = v_2A - \frac{c_2}{2}u_2^2 + \mu_1(\frac{\lambda_1}{c_1} + u_2 - \delta A) + \mu_2[(r+\delta)\lambda_1 - (1-v_2)],$$

where μ_1 and μ_2 are the costates corresponding to A and λ_1, respectively.

The necessary optimality conditions are as follows:

$$u_2 = \arg\max_{u_2 \geq 0} H \Rightarrow u_2 = \begin{cases} \mu_1/c_2 & \text{for } \mu_1 \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

$$v_2 = \arg\max_{0 \leq v_2 \leq 1} H \Rightarrow v_2 = \begin{cases} 1 \\ \text{undefined} \\ 0 \end{cases} \text{for } \sigma \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} 0, \tag{20}$$

where $\sigma(t) = A(t) + \mu_2(t)$ is the switching function.

The adjoint equations and the transversality conditions are

$$\dot{\mu}_1 = r\mu_1 - \partial H/\partial A = (r+\delta)\mu_1 - v_2, \; \mu_1(T) = 0 \tag{21}$$

$$\dot{\mu}_2 = r\mu_2 - \partial H/\partial \lambda_1 = -\delta\mu_2 - \frac{\mu_1}{c_1}, \qquad \mu_2(0) = 0. \tag{22}$$

4.2 Characterization of Possible Regimes

Corresponding to (20) there are three possible regimes which are characterized by the following lemmata.

Lemma 1. A singular solution $0 < v_2 < 1$ can only occur, if

$$c_1 < c_2. \tag{23}$$

In this case v_2 is given by $v_2 = c_2/(2c_2 - c_1)$.

Proof. Differentiation of the switching function $\sigma(t)$ with respect to t yields

$$\dot{\sigma} = \frac{\lambda_1}{c_1} + u_2 - \delta A - \delta\mu_2 - \frac{\mu_1}{c_1}$$

$$= \frac{\lambda_1}{c_1} + \mu_1\left(\frac{1}{c_2} - \frac{1}{c_1}\right) - \delta\sigma, \tag{24}$$

Differentiating a second time we obtain

$$\ddot{\sigma} = \frac{r+\delta}{c_1}\lambda_1 - \frac{1}{c_1} + v_2\left(\frac{2}{c_1} - \frac{1}{c_2}\right) + (r+\delta)\mu_1\left(\frac{1}{c_2} - \frac{1}{c_1}\right) - \delta\dot{\sigma}.$$

Substituting λ_1/c_1 from (24) and rearranging we get

$$\ddot{\sigma} = r\dot{\sigma} + \delta(r+\delta)\sigma - \frac{1}{c_1} + v_2\left(\frac{2}{c_1} - \frac{1}{c_2}\right). \tag{25}$$

Let (τ_1,τ_2) be a singular interval. i.e.

$$\sigma(t) = \dot{\sigma}(t) = \ddot{\sigma}(t) = 0 \quad \text{for} \quad t\epsilon(\tau_1,\tau_2).$$

Then from (25) follows

$$v_2 = \frac{c_2}{2c_2-c_1}. \tag{26}$$

It is easily checked that

$$0 < v_2 < 1 \quad \text{if and only if} \quad c_1 < c_2.$$

□

Lemma 2. It holds that

$$v_2(t) > 0 \quad \text{for all} \quad t\epsilon[0,T].$$

Moreover $A_o > 0$ implies $v_2(0) = 1$.

Proof. Assume the contrary, i.e. $v_2 = 0$ on some interval (τ_1',τ_2'). Define

$$\tau_1 = \inf\{\tau|\tau\epsilon[0,\tau_1'],\ v_2(t) = 0 \quad \text{for all} \quad t\epsilon[\tau,\tau_1']\}.$$

Analogously let

$$\tau_2 = \sup\{\tau|\tau\epsilon[\tau_2',T],\ v_2(t) = 0 \quad \text{for all} \quad t\epsilon[\tau_2',\tau]\}.$$

Thus, $[\tau_1,\tau_2]$ is a maximal interval including $[\tau_1',\tau_2']$ where $v_2 = 0$.

(20) implies

$$\sigma(t) \leq 0 \quad \text{for} \quad t\epsilon[\tau_1,\tau_2]. \tag{27}$$

Moreover it holds that $\sigma(\tau_1) = 0$. In the case $\tau_1 > 0$ this follows from the maximality of $[\tau_1,\tau_2]$ and the continuity of the switching function σ. For $\tau_1 = 0$ we have $\sigma(\tau_1) = A(0) + \mu_2(0) = A_0 \geq 0$ because of the transversality condition in (22). This together with (27) yields

$$\sigma(\tau_1) = 0. \tag{28}$$

Hence we conclude that

$$\dot{\sigma}(\tau_1) \leq 0. \tag{29}$$

From (25) follows

$$\ddot{\sigma}\big|_{\dot{\sigma}=0} = \delta(r+\delta)\sigma - 1/c_1 < 0 \tag{30}$$

in $[\tau_1,\tau_2]$, since $v_2 = 0$.

(30) and (29) yield

$$\dot{\sigma}(t) < 0 \quad \text{for} \quad t\varepsilon(\tau_1,\tau_2]. \tag{31}$$

This together with (28) implies $\sigma(t) < 0$ for $t\varepsilon(\tau_1,\tau_2]$. Thus, $\tau_2 = T$.

Using the boundary conditions in (18c) and (21) and (24) we get

$$\dot{\sigma}(T) = \frac{\lambda_1(T)}{c_1} + \mu_1(T)\left(\frac{1}{c_2} - \frac{1}{c_1}\right) - \delta\sigma(T) = -\delta\sigma(T).$$

This is, however, a contradiction to (27) and (31).

Finally it is easily seen by (20) and $\sigma(0) = A(0)$ that $v_2(0) = 1$ for $A_0 > 0$. □

Lemma 3. Let $c_1 < c_2$ and assume that $v_2 = 1$ in some interval $[\tau_1',\tau_2']$. Then there exists $\tau_2 \geq \tau_2'$ such that

$$v_2(t) = 1 \quad \text{for all} \quad t\varepsilon[0,\tau_2]. \tag{32}$$

Proof. In the same way as in the proof of Lemma 2 we construct a maximal interval $[\tau_1,\tau_2]$ including $[\tau_1',\tau_2']$ with $v_2 = 1$. It remains to show that $\tau_1 = 0$.

Assuming the contrary we conclude that $\sigma(\tau_1) = 0$, $\dot{\sigma}(\tau_1) \geq 0$. From (25) follows

$$\ddot{\sigma}\big|_{\dot{\sigma}=0} = \delta(r+\delta)\sigma + 1/c_1 - 1/c_2 > 0. \tag{33}$$

Here we used $\sigma \geq 0$ and assumption $c_1 < c_2$.

By a similar argument as in the proof of Lemma 2 we can show that

$$\dot{\sigma}(t) > 0, \ \sigma(t) > 0 \quad \text{for} \quad t\varepsilon(\tau_1,\tau_2] \tag{34}$$

and $\tau_2 = T$.

This leads again to $\dot{\sigma}(T) = -\delta\sigma(T)$, which contradicts (34). □

4.3 Explicit Solution

Now we are able to calculate the Stackelberg solution of (3).

Proposition 2. The Stackelberg game (3) with player 2 as leader and player 1 as follower has the following solution.

Case 1: $c_1 \geq c_2$

$$u_1(t) = 0 \tag{16}$$

$$u_2(t) = \frac{1}{c_2(r+\delta)}(1-e^{-(r+\delta)(T-t)}) \tag{17}$$

$$v_2(t) = 1. \tag{15}$$

Case 2: $c_1 < c_2$. Let τ be the switching time given by the unique nonnegative solution of

$$\left(\frac{1}{c_1} - \frac{1}{c_2}\right)\left[\frac{e^{\delta\tau}}{\delta(r+2\delta)} + \frac{e^{-(r+\delta)\tau}}{(r+\delta)(r+2\delta)} - \frac{1}{\delta(r+\delta)}\right] = A_o. \tag{35}$$

If $\tau\varepsilon[0,T)$ then the Stackelberg solution is given as

$$u_1(t) = \begin{cases} \frac{c_2-c_1}{(r+\delta)c_1(2c_2-c_1)}(e^{-(r+\delta)\tau}-e^{-(r+\delta)T})e^{(r+\delta)t} & \text{for } t\varepsilon[0,\tau) \\ \frac{c_2-c_1}{(r+\delta)c_1(2c_2-c_1)}(1-e^{-(r+\delta)(T-t)}) & \text{for } t\varepsilon(\tau,T] \end{cases} \tag{36}$$

$$u_2(t) = \begin{cases} \frac{1}{c_2(r+\delta)} + \left[\frac{1}{(r+\delta)(2c_2-c_1)}(e^{-(r+\delta)\tau}-e^{-(r+\delta)T}) - \frac{e^{-(r+\delta)\tau}}{c_2(r+\delta)}\right]e^{(r+\delta)t} & \text{for } t\varepsilon[0,\tau) \\ \frac{1}{(r+\delta)(2c_2-c_1)}(1-e^{-(r+\delta)(T-t)}) & \text{for } t\varepsilon(\tau,T] \end{cases} \tag{37}$$

$$v_2(t) = \begin{cases} 1 & \text{for } t\varepsilon[0,\tau) \\ c_2/(2c_2-c_1) & \text{for } t\varepsilon(\tau,T]. \end{cases} \tag{38}$$

If $\tau \geq T$, then the solution is the same as in case 1.

Proof. In case 1 we know from Lemma 1 that there is no singular solution and from Lemma 2 that $v_2 > 0$. According to (20) the only remaining possibility is $v_2 = 1$. As in the Nash case (Proposition 1) the controls u_1 and u_2 can be calculated.

In <u>case 2</u> from Lemma 2 and 3 follows that there are three possibilities for the allocation rate v_2:

$$v_2 = 1 \qquad \text{for} \quad t\varepsilon[0,T] \tag{39}$$

$$v_2 = \begin{cases} 1 & \text{for} \quad t\varepsilon[0,\tau) \\ c_2/(2c_2-c_1) & \text{for} \quad t\varepsilon(\tau,T] \end{cases} \tag{40}$$

$$v_2 = c_2/(2c_2-c_1) \qquad \text{for} \quad t\varepsilon[0,T]. \tag{41}$$

Note that (39) and (41), respectively, can be regarded as special cases of (40) with $\tau = T$ and $\tau = 0$, respectively.

By substituting v_2 into (18c) and (21) and using (19) and (4) the formulas (36) and (37) can be verified.

A necessary condition for the existence of a switching time τ can be deduced as follows. Solving the linear differential equation of second order (25) for $v_2 = 1$ yields

$$\sigma(t) = \frac{1}{\delta(r+\delta)}\left(\frac{1}{c_2} - \frac{1}{c_1}\right) + Ce^{-\delta t} + De^{(r+\delta)t}. \tag{42}$$

Assume that $0 < \tau < T$. Then the constants C and D can be determined by the boundary conditions

$$\sigma(0) = A_0, \; \sigma(\tau) = 0. \tag{43}$$

Moreover, since $\dot{\sigma}$ is continuous according to (24), a third condition,

$$\dot{\sigma}(\tau) = 0, \tag{44}$$

has to be satisfied. The conditions (43), (44) can be explicitely written in the form

$$C + D + \frac{1}{\delta(r+\delta)}\left(\frac{1}{c_2} - \frac{1}{c_1}\right) - A_0 = 0$$

$$Ce^{-\delta\tau} + De^{(r+\delta)\tau} + \frac{1}{\delta(r+\delta)}\left(\frac{1}{c_2} - \frac{1}{c_1}\right) = 0$$

$$C\delta e^{-\delta\tau} - D(r+\delta)e^{(r+\delta)\tau} = 0.$$

Solving any two equations of this system for C and D and substituting the solution in the third one yields equation (35) for the switching point τ.

Denoting the left hand side of (35) by $\phi(\tau)$ it turns out that

$$\phi(0) = 0$$

$$\phi'(\tau) = \frac{e^{-(r+\delta)\tau}}{r+2\delta}(e^{r\tau}-1) > 0$$

$$\lim_{\tau \to \infty} \phi(\tau) = \infty.$$

Hence $\phi(\tau)$ is a one-to-one function from $[0,\infty)$ to $[0,\infty)$, and equation (35) admits a unique solution.

If $\phi(\tau) = A_0$ has no solution $\tau \varepsilon [0,T)$, then (40) and (41) cannot be valid. Hence $v_2 = 1$ which leads to (16) and (17). □

Fig. 1 illustrates the solution in case 2.

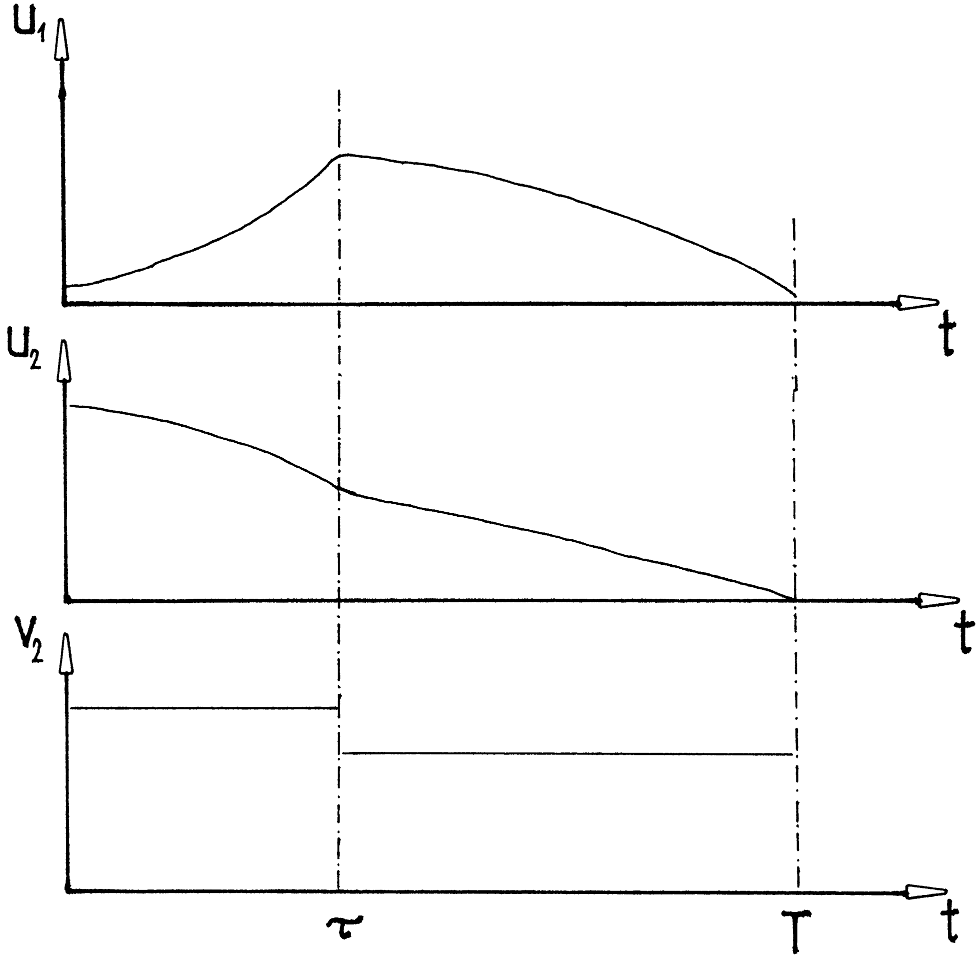


Figure 1: Optimal control trajectories for $c_1 < c_2$ and $\tau \varepsilon [0,T)$.

To interpret the shape of the Stackelberg solution it should be noted that for $A_0 = 0$ the switching point is $\tau = 0$. This means that the landlord allocates to the farmer a constant portion of the current cultivated acreage. Moreover, the switching time τ is a monotonous function of A_0. Hence, for $A_0 > 0$ the landlord reserves the whole acreage to himself in an initial interval, whereas in the 'long run' he gives a constant share of the cultivated area to the farmer.

This policy is an incentive for the farmer to cultivate in an initial interval at an increasing rate u_1, although he does not earn anything at that time. Since the farmer knows that his share 1-v will not increase after the switching point τ, his activity decreases thereafter.

It is interesting that the landlord's 'egoistic' period is at the beginning of the planning period. In analogy to maintenance problems one might have argued that the allocation v = 1 occurs at the end of the decision interval. However, this would imply that the farmer ceases to cultivate before T. Proposition 2 shows that this policy is disadvantageous for the landlord.

5. Comparison of the Nash and Stackelberg Equilibria

In this section we compare the optimal controls of the players for the Nash and the Stackelberg solution concept.

Proposition 3. The growth rate of the agricultural area in the Stackelberg case is at least as high as in the Nash case.

Proof. A simple calculation shows that for $c_1 < c_2$ the cultivation rate $u_1 + u_2$ given by (36), (37) exceeds the corresponding sum given by (16), (17). For $c_1 \geq c_2$ the solutions coincide which completes the proof. □

According to Proposition 1 the farmer earns nothing in the Nash game. Since by choosing $u_1 = 0$ he can always get a nonnegative payoff, as a Stackelberg follower he will not be worse off than as a Nash player. The same holds true for the landlord.For open-loop information structure it is generally true that the leader's payoff in the Stackelberg game is at least as high as in the Nash game.

This shows that both players prefer to play Stackelberg rather than Nash. This property is known as Stackelberg dominance.

It can be shown that in a Stackelberg game with the farmer as leader and the landlord as follower the solution coincides with the Nash equilibrium (15) - (17). The sharecropping game is stable in the sense that given the condition to play Stackelberg both players do not prefer to play the same role. Note that this property is opposite to the situation in the capitalism game, where workers as well as capitalists prefer to act as followers (see Pohjola, 1984).

We now compare the control trajectories for both solution concepts. The only interesting case is case 2 of Proposition 2 where switching occurs, i.e. $\tau \epsilon [0,T)$.

Denoting by u_i^N and u_i^S the Nash solution (16), (17) and the Stackelberg equilibrium (36), (37), respectively, the following proposition is valid.

Proposition 4. It holds that

$$0 = u_1^N(t) < u_1^S(t) \quad \text{for} \quad t \epsilon [0,T] \tag{45}$$

$$u_2^N(t) > u_2^S(t) \quad \text{for} \quad t \epsilon [0,T]. \tag{46}$$

The proof is obvious.

6. Pareto Solutions, Optimal Threats, and Cooperative Nash Equilibria

6.1 Pareto Solutions

We now consider a cooperative principal-agent situation. Pareto-optimal solutions of the sharecropping game (3) can be found by maximizing the weighted criterion

$$J = J_1 + \mu J_2 \tag{47}$$

for J_1 and J_2 given in (2) and for different values of μ such that $0 < \mu < \infty$. Here μ is given and measures the bargaining power of the landlord. For $\mu > 1$ its relative importance is higher than that of the farmer.

The following result characterizes the Pareto solution depending on μ.

Proposition 5. The game (47), (3bcd) has the following Pareto solution:

For $\mu < 1$

$$u_1 = \frac{1}{c_1(r+\delta)}(1-e^{-(r+\delta)(T-t)}) \tag{48a}$$

$$u_2 = \frac{1}{\mu c_2(r+\delta)}(1-e^{-(r+\delta)(T-t)}) \tag{48b}$$

$$v_2 = 0. \tag{48c}$$

For $\mu > 1$

$$u_1 = \frac{\mu}{c_1(r+\delta)}(1-e^{-(r+\delta)(T-t)}) \tag{49a}$$

$$u_2 = \frac{1}{c_2(r+\delta)}(1-e^{-(r+\delta)(T-t)}) \tag{49b}$$

$$v_2 = 1. \tag{49c}$$

For $\mu = 1$ v_2 is arbitrary in $[0,1]$, whereas u_1 and u_2 are given by (48a) and (49b), respectively.

Proof. The Pareto game leads to an ordinary optimal control problem with the Hamiltonian

$$H = (1-v_2)A - \frac{c_1}{2}u_1^2 + \mu(v_2A - \frac{c_2}{2}u_2^2) + \lambda(u_1+u_2-\delta A).$$

Necessary optimality conditions are

$$u_1 = \arg\max_{u_1 \geq 0} H \Rightarrow u_1 = \begin{cases} \lambda/c_1 & \text{for } \lambda \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{50}$$

$$u_2 = \arg\max_{u_2 \geq 0} H \Rightarrow u_2 = \begin{cases} \lambda/(c_2\mu) & \lambda \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{51}$$

$$v_2 = \arg\max_{0 \leq v_2 \leq 1} H \Rightarrow v_2 = \begin{cases} 0 \\ \text{undefined for} \\ 1 \end{cases} A(\mu-1) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} 0 \tag{52}$$

$$\dot{\lambda} = r\lambda - \partial H/\partial A = (r+\delta)\lambda - 1 - \mu v_2 + v_2 \tag{53}$$

$$\lambda(T) = 0. \tag{54}$$

From (53) and (54) follows

$$\lambda(t) = \int_t^T e^{-(r+\delta)(\tau-t)}(1+\mu v_2-v_2)d\tau > 0, \tag{55}$$

where the positivity of λ results from $\mu > 0$, $v_2 \varepsilon [0,1]$. According to (50), (51) we get $u_1 > 0$, $u_2 > 0$, and thus $A > 0$. Hence, (52) may be written as

$$v_2 = \begin{cases} 0 \\ \text{undefined} \quad \text{for} \\ 1 \end{cases} \mu \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} 1. \tag{56}$$

Substituting (56) into (55), evaluating the integral and using (50) and (51) leads to (48) and (49). □

Note that there is a Pareto equilibrium for every $\mu > 0$. The bargaining power μ is determined according to the procedure outlined in Section 1. To this end let us derive optimal threats announced by the players to affect the negotiated solution to their own advantage.

6.2 Optimal Threats and Bargaining Solution

Liu (1973) gives the following sufficient conditions for the existence of optimal threat strategies.

Denote by (u_1^P, u_2^P, v_2^P) the Pareto-optimal solution and by (u_1^*, u_2^*, v_2^*) the optimal threat strategies. Then it is sufficient to find a constant $0 < \mu < \infty$ such that the following conditions hold:

$$J_1(u_1^P, u_2^P, v_2^P) + \mu J_2(u_1^P, u_2^P, v_2^P) = \max_{u_1, u_2, v_2} (J_1 + \mu J_2) \tag{57}$$

$$J_1(u_1^*, u_2^*, v_2^*) - \mu J_2(u_1^*, u_2^*, v_2^*) = \min_{u_2, v_2} \max_{u_1} (J_1 - \mu J_2) = \max_{u_1} \min_{u_2, v_2} (J_1 - \mu J_2) \tag{58}$$

$$J_1(u_1^P, u_2^P, v_2^P) - \mu J_2(u_1^P, u_2^P, v_2^P) = J_1(u_1^*, u_2^*, v_2^*) - \mu J_2(u_1^*, u_2^*, v_2^*). \tag{59}$$

Note that (58) means that (u_1^*, u_2^*, v_2^*) is a saddle-point of $J_1 - \mu J_2$.

To apply Liu's theorem we first calculate the optimal threat strategies by solving (58).

Lemma 4. The optimal threat strategies of the sharecropping game are given by the Nash solution (15) - (17)

$$u_1^* = u_1^N, \; u_2^* = u_2^N, \; v_2^* = v_2^N. \tag{60}$$

Proof. To solve the saddle-point problem (58) we define the Hamiltonian

$$H = (1-v_2)A - \frac{c_1}{2} u_1^2 - \mu(v_2 A - \frac{c_2}{2} u_2^2) + \lambda(u_1 + u_2 - \delta A).$$

The necessary optimality conditions are

$$u_1^* = \arg\max_{u_1 \geq 0} H \Rightarrow u_1^* = \begin{cases} \lambda/c_1 & \text{for } \lambda \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{61}$$

$$u_2^* = \arg\min_{u_2 \geq 0} H \Rightarrow u_2^* = \begin{cases} -\lambda/(\mu c_2) & \text{for } \lambda \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{62}$$

$$v_2^* = \arg\min_{0 \leq v_2 \leq 1} H \Rightarrow v_2 = \begin{cases} 0 \\ \text{undefined} \\ 1 \end{cases} \text{ for } A \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} 0 \tag{63}$$

$$\dot{\lambda} = (r+\delta)\lambda - 1 + v_2 + \mu v_2, \; \lambda(T) = 0. \tag{64}$$

Since from $A_0 > 0$ follows $A(t) > 0$ we conclude from (63) that $v_2 = 1$. Substituting this into (64) yields

$$\lambda(t) = -\frac{\mu}{r+\delta}(1-e^{-(r+\delta)(T-t)}) < 0. \tag{65}$$

Using (61) and (62) yields (60). □

We now evaluate the objective functionals for Pareto solutions and threat strategies to get the functions

$$K^P(\mu) = J_1(u_1^P,u_2^P,v_2^P) - \mu J_2(u_1^P,u_2^P,v_2^P) \tag{66}$$

$$K^*(\mu) = J_1(u_1^*,u_2^*,v_2^*) - \mu J_2(u_1^*,u_2^*,v_2^*). \tag{67}$$

Lemma 5. It holds that

$$K^P(\mu) = \begin{cases} M + N\left(\frac{1}{c_1} + \frac{3}{\mu c_2}\right) & \text{for} \quad \mu < 1 \quad (68a) \\ -\mu\left[M + N\left(\frac{1}{c_2} + \frac{3\mu}{c_1}\right)\right] & \text{for} \quad \mu > 1 \quad (68b) \end{cases}$$

$$K^*(\mu) = -\mu(M+N/c_2), \tag{68c}$$

where M and N are constants given by

$$M = \frac{A_0}{r+\delta}\left[1-e^{-(r+\delta)T}\right] \tag{69}$$

$$N = \frac{1}{r+\delta}\left[-\frac{r+\delta}{r\delta(r+2\delta)}e^{-rT} + \frac{1}{\delta(r+\delta)}e^{-(r+\delta)T} - \right.$$

$$\left. - \frac{1}{2(r+\delta)(r+2\delta)}e^{-2(r+\delta)T} + \frac{1}{2r(r+\delta)}\right] \tag{70}$$

The proof is given in the Appendix.

The remaining task in proving the existence of the Nash bargaining equilibrium is to find μ such that (59) holds, i.e., such that

$$K^P(\mu) = K^*(\mu),$$

where $K^P(\mu)$ and $K^*(\mu)$ are given by Lemma 5.

Proposition 6. The curves $K^P(\mu)$ and $K^*(\mu)$ given by (68) have a unique intersection at $\mu = 1$.

Proof. Obviously, M is positive. N may be written as

$$N = \frac{1}{r+\delta}\psi_1(T),$$

where ψ_1 denotes the expression in square brackets in (70). It holds that $\psi_1(0) = 0$. To show $N > 0$ is suffices to show that $\psi_1'(T) > 0$. We observe that

$$\psi_1'(T) = e^{-rT}\psi_2(T) \tag{71}$$

with

$$\psi_2(T) = \frac{r+\delta}{\delta(r+2\delta)} - \frac{1}{\delta}e^{-\delta T} + \frac{1}{r+2\delta}e^{-(r+2\delta)T}.$$

It can easily be checked that

$$\psi_2(0) = 0, \ \psi_2'(T) > 0$$

which shows $\psi_2(T) > 0$, and, by (71), $\psi_1'(T) > 0$.

The shape of the curves $K^P(\mu)$ and $K^*(\mu)$ is depicted in Fig. 2.

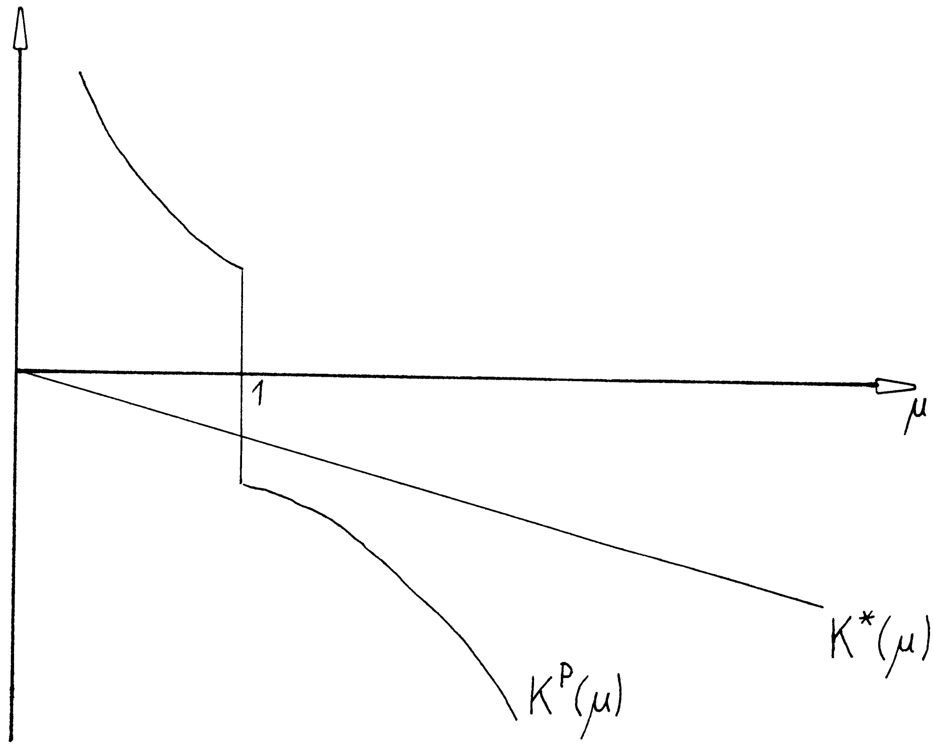


Figure 2: The curves $K^P(\mu)$ and $K^*(\mu)$.

It should be noted that the curve $K^P(\mu)$ is continuous at $\mu = 1$, since each value between the one-sided limits can be realized by proper choice of v_2. This is possible, since according to Proposition 5 in the case $\mu = 1$ v_2 is arbitrary. From Fig. 2 the existence and the uniqueness are evident. □

As unexpected result we obtain that the bargaining power is the same for each player.

7. Concluding Remarks

The purpose of the present paper is twofold. First, it is an attempt to discuss a principal-agent situation with particular emphasis on the intertemporal character of that problem. Second, it provides an example of a differential game which is simple enough to be explicitly solvable for different assumption on information. Similar to the capitalism game of Lancaster (1973) the Stackelberg solution differs from the noncooperative Nash equilibrium. For simplicity we have restricted our analysis to open-loop Nash and Stackelberg equilibria.

The model can be extended by including a rent payed by the farmer to the landlord. Moreover it is possible to take into consideration that the farmer earns a wage for his cultivation activities.

The game allows for other interpretation, e.g. as maintenance model. The principal (player 2) ownes a machine the output of which depends on its quality A. The quality can be improved by maintenance measures u_1 and u_2, respectively. The agent (player 1) gets rights of usufruct for his maintenance activities. The principial allocates the machine's output with rate $1-v_2$ to his agent.

Another interpretation is of more academic interest. If concerns the well-known relationship between a professor and his assistant. The variable A may denote the value of a research project. The professor has the power to transfer a portion of the remuneration obtained for the project to the assistant.

Appendix

Proof of Lemma 5. For the proof we need the following two obvious technical results.

First, for

$$u_1 = B_1(1-e^{-(r+\delta)(T-t)}),\ u_2 = B_2(1-e^{-(r+\delta)(T-t)}),\ v_2 = B_3 \qquad (A.1)$$

the solution of the state equation (1) is given by

$$A = G_1 + G_2 e^{(r+\delta)t} + G_3 e^{-\delta t}, \qquad (A.2)$$

where

$$G_1 = \frac{B_1+B_2}{\delta},\ G_2 = -\frac{B_1+B_2}{r+2\delta}e^{-(r+\delta)T},\ G_3 = A_o - G_1 - G_2. \tag{A.3}$$

Second, for (A.1) and (A.2) the integrands in the objective functionals (2) are given by:

$$D_1e^{-rt} + D_2e^{\delta t} + D_3e^{-(r+\delta)t} + D_4e^{(r+2\delta)t} \tag{A.4}$$

where for <u>player 1</u>

$$D_1 = (1-B_3)G_1 - (c_1/2)B_1^2,\quad D_2 = (1-B_3)G_2 + c_1B_1^2e^{-(r+\delta)T},$$
$$D_3 = (1-B_3)G_3,\quad D_4 = -(c_1/2)B_1^2e^{-2(r+\delta)T}, \tag{A.5}$$

and for <u>player 2</u>

$$D_1 = B_3G_1 - (c_2/2)B_2^2,\quad D_2 = B_3G_2 + c_2B_2^2e^{-(r+\delta)T},$$
$$D_3 = B_3G_3,\quad D_4 = -(c_2/2)B_2^2e^{-2(r+\delta)T}. \tag{A.6}$$

We are now ready to prove Lemma 5.

<u>Case 1: Pareto $\mu < 1$</u>:

First we evaluate the objective functional $J_1(u_1^P,u_2^P,v_2^P)$ for <u>player 1</u> in the case $\mu < 1$. The constants B_1,B_2,B_3 in (A.1) are given by

$$B_1 = \frac{1}{c_1(r+\delta)},\ B_2 = \frac{1}{\mu c_2(r+\delta)},\ B_3 = 0.$$

Using (A.9) and (A.11) we get

$$D_1 = \frac{1}{r+\delta}\left(\frac{1}{c_1\delta} + \frac{1}{\mu c_2\delta} - \frac{1}{2c_1(r+\delta)}\right)$$

$$D_2 = -\frac{e^{-(r+\delta)T}}{r+\delta}\left(\frac{1}{c_1(r+2\delta)} + \frac{1}{\mu c_2(r+2\delta)} - \frac{1}{c_1(r+\delta)}\right)$$

$$D_3 = A_o - \frac{1}{c_1\delta(r+\delta)} - \frac{1}{\mu c_2\delta(r+\delta)} + \frac{e^{-(r+\delta)T}}{(r+\delta)(r+2\delta)}\left(\frac{1}{c_1} + \frac{1}{\mu c_2}\right)$$

$$D_4 = -\frac{e^{-2(r+\delta)T}}{2c_1(r+\delta)^2}.$$

Integrating (A.4) in [0,T] we obtain

$$J_1(u_1^P,u_2^P,v_2^P) = e^{-rT}\Big[-\frac{1}{r(r+\delta)}\Big(\frac{1}{c_1\delta}+\frac{1}{\mu c_2\delta}-\frac{1}{2c_1(r+\delta)^2}\Big)-\frac{1}{\delta(r+\delta)}\Big(\frac{1}{c_1(r+2\delta)}+$$

$$+\frac{1}{\mu c_2(r+2\delta)}-\frac{1}{c_1(r+\delta)}\Big)-\frac{1}{2c_1(r+\delta)^2(r+2\delta)}\Big]+$$

$$e^{-(r+\delta)T}\Big[-\frac{1}{(r+\delta)}\Big(A_0-\frac{1}{\delta(r+\delta)c_1}-\frac{1}{\mu c_2\delta(r+\delta)}\Big)+\frac{1}{\delta(r+\delta)}\Big(\frac{1}{c_1(r+2\delta)}+$$

$$+\frac{1}{\mu c_2(r+2\delta)}-\frac{1}{c_1(r+\delta)}\Big)+\frac{1}{(r+\delta)^2(r+2\delta)}\Big(\frac{1}{c_1}+\frac{1}{\mu c_2}\Big)\Big]+$$

$$+\Big[\frac{1}{r(r+\delta)}\Big(\frac{1}{c_1\delta}+\frac{1}{\mu c_2\delta}-\frac{1}{2c_1(r+\delta)}\Big)+\frac{1}{r+\delta}\Big(A_0-\frac{1}{c_1\delta(r+\delta)}-\frac{1}{\mu c_2\delta(r+\delta)}\Big)\Big]+$$

$$+e^{-2(r+\delta)T}\Big[-\frac{1}{(r+\delta)^2(r+2\delta)}\Big(\frac{1}{c_1}+\frac{1}{\mu c_2}\Big)+\frac{1}{2c_1(r+2\delta)(r+\delta)^2}\Big]. \qquad \text{(A.7)}$$

For player 2 we get from (A.6)

$$D_1 = \frac{1}{2\mu^2c_2(r+\delta)^2},\quad D_2 = \frac{e^{-(r+\delta)T}}{\mu^2c_2(r+\delta)^2}$$

$$D_3 = 0,\quad D_4 = -\frac{e^{-2(r+\delta)T}}{2\mu^2c_2(r+\delta)^2}.$$

This yields

$$J_2(u_1^P,u_2^P,v_2^P) = e^{-rT}\Big[\frac{1}{2\mu^2c_2(r+\delta)^2r}+\frac{1}{\mu^2c_2\delta(r+\delta)^2}-\frac{1}{2\mu^2c_2(r+\delta)^2(r+2\delta)}\Big]-$$

$$-\frac{1}{2\mu^2c_2(r+\delta)^2r}-\frac{e^{-(r+\delta)T}}{\mu^2c_2\delta(r+\delta)^2}+\frac{e^{-2(r+\delta)T}}{2\mu^2c_2(r+\delta)^2(r+2\delta)}. \qquad \text{(A.8)}$$

Using (66), (A.7) and (A.8) a straightforward calculation shows (68a).

Case 2: Pareto $\mu > 1$:

The procedure is analogous to the case $\mu > 1$. The constants B_1, B_2, B_3 are as follows:

$$B_1 = \frac{\mu}{c_1(r+\delta)},\quad B_2 = \frac{1}{c_2(r+\delta)},\quad B_3 = 1.$$

Using (A.3) and (A.5) yields for player 1

$$D_1 = -\frac{\mu^2}{2c_1(r+\delta)},\quad D_2 = \frac{\mu^2e^{-(r+\delta)T}}{c_1(r+\delta)^2}$$

$$D_3 = 0,\ D_4 = -\frac{\mu^2 e^{-2(r+\delta)T}}{2c_1(r+\delta)^2}.$$

Integrating (A.4) in [0,T] we get

$$J_1(u_1^P,u_2^P,v_2^P) = e^{-rT}\left[\frac{\mu^2}{2c_1 r(r+\delta)^2} + \frac{\mu^2}{c_1\delta(r+\delta)^2} - \frac{\mu^2}{2c_1(r+\delta)^2(r+2\delta)}\right] -$$

$$- \frac{\mu^2}{2c_1 r(r+\delta)^2} - \frac{\mu^2 e^{-(r+\delta)T}}{c_1\delta(r+\delta)^2} + \frac{\mu^2 e^{-2(r+\delta)T}}{2c_1(r+\delta)^2(r+2\delta)}. \tag{A.9}$$

For <u>player 2</u> we obtain from (A.6)

$$D_1 = \frac{1}{r+\delta}\left(\frac{\mu}{c_1\delta} + \frac{1}{c_2\delta} - \frac{1}{2c_2(r+\delta)}\right)$$

$$D_2 = -\frac{e^{-(r+\delta)T}}{r+\delta}\left(\frac{\mu}{c_1(r+2\delta)} + \frac{1}{c_2(r+2\delta)} - \frac{1}{c_2(r+\delta)}\right)$$

$$D_3 = A_0 - \frac{\mu}{c_1(r+\delta)\delta} - \frac{1}{c_2\delta(r+\delta)} + \frac{e^{-(r+\delta)T}}{(r+\delta)(r+2\delta)}\left(\frac{\mu}{c_1} + \frac{1}{c_2}\right)$$

$$D_4 = -\frac{e^{-2(r+\delta)T}}{2c_2(r+\delta)^2}.$$

This yields

$$J_2(u_1^P,u_2^P,v_2^P) = e^{-rT}\left[-\frac{\mu}{c_1\delta r(r+\delta)} - \frac{1}{c_2\delta r(r+\delta)} + \frac{1}{2c_2 r(r+\delta)^2} - \frac{\mu}{c_1\delta(r+\delta)(r+2\delta)} -\right.$$

$$\left. - \frac{1}{c_2\delta(r+\delta)(r+2\delta)} + \frac{1}{c_2\delta(r+\delta)^2} - \frac{1}{2c_2(r+\delta)^2(r+2\delta)}\right] +$$

$$+e^{-(r+\delta)T}\left[-\frac{1}{r+\delta}\left(A_0 - \frac{\mu}{c_1(r+\delta)\delta} - \frac{1}{c_2\delta(r+\delta)}\right) + \frac{\mu}{c_1\delta(r+\delta)(r+2\delta)} +\right.$$

$$\left. + \frac{1}{c_2\delta(r+\delta)(r+2\delta)} - \frac{1}{c_2\delta(r+\delta)^2} + \frac{1}{r+\delta}\left(\frac{\mu}{c_1(r+\delta)(r+2\delta)} + \frac{1}{c_2(r+\delta)(r+2\delta)}\right)\right] +$$

$$+e^{-2(r+\delta)T}\left[-\frac{\mu}{c_1(r+\delta)^2(r+2\delta)} - \frac{1}{c_2(r+\delta)^2(r+2\delta)} + \frac{1}{2c_2(r+\delta)^2(r+2\delta)}\right] +$$

$$+\left[\frac{\mu}{c_1 r\delta(r+\delta)} + \frac{1}{c_2\delta r(r+\delta)} - \frac{1}{2c_2 r(r+\delta)^2} + \frac{1}{r+\delta}\left(A_0 - \frac{\mu}{c_1(r+\delta)\delta} - \frac{1}{c_2\delta(r+\delta)}\right)\right]. \tag{A. 10}$$

Using (66), (A.9) and (A.10) it is easy to verify (68b).

Case 3: Threat strategies

Proceeding analogously as before we obtain

$$B_1 = 0,\ B_2 = \frac{1}{c_2(r+\delta)},\ B_3 = 1.$$

Player 1:

$$D_1 = D_2 = D_3 = D_4 = 0.$$

Player 2:

$$D_1 = \frac{1}{c_2(r+\delta)}\left(\frac{1}{\delta} - \frac{1}{2(r+\delta)}\right)$$

$$D_2 = \frac{e^{-(r+\delta)T}}{c_2(r+\delta)}\left(\frac{1}{r+\delta} - \frac{1}{r+2\delta}\right)$$

$$D_3 = A_0 - \frac{1}{c_2\delta(r+\delta)} + \frac{e^{-(r+\delta)T}}{c_2(r+\delta)(r+2\delta)}$$

$$D_4 = -\frac{e^{-2(r+\delta)T}}{2c_2(r+\delta)^2}.$$

This yields

$$J_1(u_1^*,u_2^*,v_2^*) = 0$$

$$\begin{aligned} J_2(u_1^*,u_2^*,v_2^*) = {} & e^{-rT}\left[\frac{1}{c_2\delta r(r+\delta)} + \frac{1}{2c_2r(r+\delta)^2} - \frac{1}{c_2\delta(r+\delta)(r+2\delta)} + \frac{1}{c_2\delta(r+\delta)^2} - \frac{1}{2c_2(r+\delta)^2(r+2\delta)}\right] + \\ & + e^{-(r+\delta)T}\left[-\frac{1}{r+\delta}\left(A_0 - \frac{1}{c_2\delta(r+\delta)}\right) + \frac{1}{c_2\delta(r+\delta)(r+2\delta)} - \frac{1}{c_2\delta(r+\delta)^2} + \frac{1}{c_2(r+\delta)^2(r+2\delta)}\right] + \\ & + e^{-2(r+\delta)T}\left[-\frac{1}{c_2(r+\delta)^2(r+2\delta)} + \frac{1}{2c_2(r+\delta)^2(r+2\delta)}\right] + \\ & + \left[\frac{1}{c_2\delta r(r+\delta)} - \frac{1}{2c_2r(r+\delta)^2} + \frac{1}{r+\delta}\left(A_0 - \frac{1}{c_2\delta(r+\delta)}\right)\right]. \end{aligned}$$

From this and (67) we obtain (68c). □

References

T. BASAR and G.J. OLSDER, "Dynamic Noncooperative Game Theory", Academic Press, London, 1982.

G. FEICHTINGER and R.F. HARTL, "Optimale Kontrolle ökonomischer Prozesse: Anwendungen des Maximumprinzips in den Wirtschaftswissenschaften", de Gruyter, Berlin, 1986.

K. LANCASTER, The dynamic inefficiency of capitalism, J. Pol. Econ. 81 (1973), 1092-1109.

P.-T. LIU, Optimal threat strategies in differential games, J. Math. Anal. Appl. 43 (1973), 161-169.

M. POHJOLA, Threats and bargaining in capitalism: a differential game view, J. Econ. Dyn. Contr. 8 (1984), 291-302.

The help of E. Dockner is gratefully acknowledged.

Section 7 Property Rights and Fairness

Managerialism versus the Property Rights Theory of the Firm

Thomas Kaulmann

Summary: The consequences for corporate returns of the separation of ownership and control in the large publicly-held corporation are evaluated on the basis of two different theoretical schools: The managerial theories predict lower returns for manager-controlled corporations compared with owner-operated enterprises. On the other hand, the property rights theory of the firm predicts that returns for the manager-controlled corporation will not be lower, since the competitive processes in the corporate environment restrict the managers' discretion and force them to operate efficiently. The validity of one standpoint or the other could not be clarified empirically until the mid-1970's, since contradictory results had been obtained up to that point. Analysis of the most recent empirical work in this field shows extensive corroboration of the property rights theory. The large publicly-held corporation can thus be seen as an efficient form of enterprise, and the property rights theory of the firm can be considered the superior theory.

Table of Contents:

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

1. Definition of the Problem

Since the early 1930's at the latest, when the work of Berle and Means[1] appeared, the large publicly-held corporation has been a frequently discussed subject of study. The two authors founded the "tradition" of "managerialists", who developed different theoretical approaches to the economic consequences of separation of ownership and control in the modern publicly-held corporation. These approaches have been tested in numerous empirical studies, the results of which - taken as a whole - have been very contradictory.

The property rights theory, developed by Alchian and Demsetz in the course of the 1960's, is a microeconomic theory with many fields of application. In the field of theory of the firm, it makes possible economic analysis of a corporation's charter. One of many relationships fixed within the legal framework of charters is that between ownership and control, so that by applying the property rights theory, assertions can be made regarding the economic consequences of the separation of ownership and control. The results of this application show that conclusions derived from the property rights theory partially contradict conclusions reached by the managerialists. This paper intends to help clarify these contradictory conclusions. To this end, the paper will begin by outlining the two controversial viewpoints once again (section 2). Section 3 provides a summary overview of the current state of empirical research; section 4 contains conclusions based on the preceding discussion.

2. Theoretical Assertions Regarding the Separation of Ownership and Control in the Modern Publicly-Held Corporation

In the "classic corporation", the entrepreneur is simultaneously both owner and highest-ranking manager. Neo-classical theory is based on exactly this unity of ownership and control, and furthermore views the corporation as a "one-man-operation" whose only goal is to realize maximum profit. In reality, corporations have a considerable number of employees, and the respective owners are not necessarily identical with the highest-ranking managerial body in the corporation. The managerialists have found these facts to be of economic relevance, and have taken them into account in their theories.

---

[1] cf. Berle/Means (1932).

2.1 Managerialists

The consequences of the separation of ownership and control, which can be empirically observed, are outlined by the managerialists as follows. The owners assign power of direction to salaried managers. Since the owners have insufficient opportunity for monitoring operations, a certain amount of discretionary power is placed in the hands of these managers. Accordingly, the managers need not completely adhere to tasks assigned by the owners, but rather have the opportunity to pursue their own goals. This pursuit of their own goals leads to deviation from the profit-maximizing behavior attributed to the corporation in neo-classical theory.

The managerialists have proposed various target functions which managers might have, and in each case have derived the corresponding corporate performance from consistent mathematical models.

The sales-maximizing firm should be mentioned here first[2]. Baumol initially justifies his assumption by referring to the negative consequences for the corporation and management resulting from decreased sales[3]. Among other points, he mentions declining profit potential, increasing scepticism on the part of banks and investors, declining product popularity and decreasing management income (since this is frequently tied to sales). These are arguments which partly support the view that higher sales figures are only strived for in order to achieve higher profits. However, Baumol is of the opinion that sales itself has become a maximization goal, as supported by actual observations:[4]

- managers always (first) mention sales figures when evaluating economic success;
- the success of managers is measured in terms of sales figures;
- managers retain unprofitable corporate divisions to prevent sales figures from decreasing.

---

[2] cf. Baumol (1967).

[3] cf. Baumol (1967)), pp. 45-52.

[4] cf. Baumol (1967), p. 46.

In Baumol's model, profit operates as a constraint which must be fulfilled to satisfy the stockholders. In a formal model, maximization of sales under this constraint using the Lagrange function results in a sales-maximizing corporation extending its sales value beyond the point at which marginal sales and marginal costs are equal. This means that the corporation operates with greater output and lower profit compared to a profit-maximizing corporation.

The dynamic equivalent of Baumol's static model is the growth-maximization model by Marris.[5] He uses specific psychological traits found among managers, and the resulting utility functions, to justify his model:[6] if the managers are already at the head of the corporation, then within the context of the organization it is only corporate growth which can continue to satisfy their intense dedication to performance and the "drive to reach the top." Marris adds another constraining variable to the utility function of managers: the evaluation ratio, which is the ratio of corporation market value to book value.[7] The evaluation ratio influences the probability of takeover by another corporation, an event which could result in management either losing their positions or being assigned to an area of responsibility in which their autonomy is reduced. For this reason, managers will strive for a certain minimum evaluation ratio in order to achieve a degree of security in terms of takeover prevention. Like Baumol, Marris can also prove in his model that the growth-maximizing corporation increases output beyond the optimal profit-maximizing level.

In his managerial model, Williamson incorporates the assumption that management attempts to satisfy personal goals regarding earnings, security, status, power, social interests and professionalism. Measures which may be well-suited for satisfying these personal goals do not contribute to corporate productivity. Among the measures which management will take, Williamson mentions expansion of staff departments, expenditures which immediately benefit managers (size and furnishing of offices, company car, chauffeur, etc.), and increasing profit beyond levels expected by the stockholders.[8] In the analytic formulation and evaluation of his model, Williamson shows that

---

5) cf. Marris (1964).

6) cf. Marris (1964), pp. 49-56.

7) cf. Marris (1964), pp. 106-107.

8) cf. Williamson (1964), pp. 34-36.

compared to owner-controlled corporations, manager-controlled corporations will operate with higher output, greater preferred expenses and lower profit margins.[9)]

This discussion of different managerial models shows that despite being based on different assumptions, and in contrast to the profit maximization model, they reach the same conclusion regarding profitability: manager-controlled corporations operate with lower profitability compared to owner-controlled corporations.

2.2 Property Rights Theory of the Firm

The property rights theory is a new microeconomic approach which also allows an economic analysis of the separation of ownership and control in the large publicly-held corporation. The property rights theory emphasizes that it is not ownership of economic resources per se that is of particular interest, but rather that it is the ownership of the rights associated with the resources which constitutes the economic value of the resources.[10)] To be able to use this theoretical foundation for analytical purposes, the rights associated with economic goods have been broken down into rights bundles:[11)]

(1) the right to use economic resources
(2) the right to modify the form and substance of the resources
(3) the right to benefit from the use of the resources
(4) the right to transfer the resources

These rights can naturally also be observed in corporations, where it is common practice to combine the first two rights into the right of coordination.[12)]

In addition to influencing the value of economic resources, the basic organization of property rights also influences the behavior of

9) cf. Williamson (1964), pp. 52-54. In more recent work Williamson has changed from a managerial position to a property rights standpoint. See Williamson (1975, 1985).

10) for basic literature on property rights theory, see: Alchian (1965); Demsetz (1967); Alchian/Demsetz (1973), and for articles presenting an overview in German, see: Leipold (1978); Tietzel (1981); Meyer (1983); Gäfgen (1984).

11) cf. for example Alchian/Demsetz (1972), p. 783.

12) cf. for example Picot (1981), pp. 161 ff.

individuals dealing with these resources. The basic organization of property rights can vary, with the result that rights bundles may be distributed among one or more individuals or that single property rights are only realized to an attenuated degree. Thus, the greater the number of individuals among which single rights bundles are distributed, and the more restricted single rights bundles are, the lesser the value of the resource involved, with the result that the individual deals with the resource in a more uneconomic manner.

Further elements of the property rights theory to be mentioned are transaction costs and maximization of individual benefit. Property rights analysis assumes that people act in their own interest, e.g. individuals attempt to increase their own welfare as far as possible and avoid welfare losses as far as possible. The assumption of maximization of individual benefit is a central element in the field of the theory of the firm: here too, the assumption of profit maximization is negated and replaced with a more comprehensive assumption of behavior.

A prerequisite for the ability to dispose freely of property rights is agreement between individuals regarding the exchange of these rights. These contracts cannot be realized costfree. The costs incurred in searching for a contractual partner, and in closing and monitoring the contract - designated as transaction costs - are explicitly taken into account in the property rights approach. Transaction costs become particularly relevant when considering, in conjunction with the assumption of maximization of individual benefit, the possibility of behavior in violation of the contract: since such behavior in violation of the contract cannot be detected without incurring costs, this type of behavior is possible until the damage it causes is greater than the expected control costs. The degree of damage is thus limited by the incurred transaction costs.[13)]

Fundamental property rights among the participating individuals of a firm are regulated in the corporation's charter. On the one hand, the charter takes into account voluntary agreements reached under private law, and on the other hand includes provisions stipulated by the government, such as codetermination and mandatory disclosure, which are not influenced by individual contracts. Thus, different charters

---

13) at this point, the close theoretical relationship between property rights theory and the principal-agent problem becomes apparent.

are associated with different levels of transaction costs, since rights can be realized and monitored to various degrees and distributed in different ways to different persons in the individual charter. Transaction costs thereby constitute a measure for evaluating the economic effects of a corporation's charter.

Without going into detail at this point regarding these types of extensive analyses[14], there follows a brief presentation of the most important results. In the classic corporation, all three rights bundles as defined above are in the hands of the owner-entrepreneur. He directly benefits from his organizational and supervisory abilities through the profits yielded and from possible capitalization of the corporation. He thus has a strong economic interest in efficient coordination of resources and supervision of corporation employees. As a result of the concentration of property rights, this form of a corporation's charter is associated with low transaction costs.

The already-mentioned separation of ownership and control occurs quite frequently in the publicly-held corporation. The corporation owners, who are often quite numerous, and who are vested with the rights of profit acquisition and the transfer of their shares, hire managers to lead the corporation. The costs of supervision of management may be relatively high, since information acquisition is costly and a coordinated organized procedure for the stockholders is also associated with high transaction costs. Individual small stockholders are rarely involved in measures to regulate management, since the associated costs usually exceed the benefits of such actions. Consequently, in an initial analysis, the wide dispersion of property rights in the publicly-held corporation leads to relatively high transaction costs for this form of corporate charter.

However, the property rights theorists present further arguments which refine the analysis of the separation of ownership and control. In particular, they argue that the disincentive effects of the separation of ownership and control are moderated by competition in external capital and product markets and the market for managerial capital.[15]

---

14) cf. for example Picot (1981); Picot/Michaelis (1984); Kaulmann (1987).

15) cf. for example Alchian (1984), pp. 43 ff.; De Alessi (1983), pp. 73 ff.; Demsetz (1983), pp. 387 ff.; Furubotn/Pejovich (1972), pp. 1149 ff.; Picot (1981), p. 167; Picot/Kaulmann (1985), pp. 958 ff.; Picot/Michaelis (1984), pp. 259 ff.

In addition to the issue of internal distribution of property rights, discussion is mainly of control through the capital market. There is a simple but effective instrument available to the owners to regulate the managers: they can dispose of their capital shares on the capital market. Potential owners acquire decision-related information on the corporation and its management and evaluate this information when making investment decisions. As a result, the capital market takes on an informational search function, and serves to evaluate management performance. These evaluations could lead to negative changes in the corporation's market value, which would also have consequences for management. Firstly, possibilities for financing (i.e. increasing external debt and equity capital) are worsened. Secondly, the probability of takeover by other institutional investors increases with a low market value. A takeover could lead to replacement of management in order to increase the earning power of the corporation.

Closely related to the effects of the capital market are the further constraints placed on managerial discretion by sales markets. Under workable competition, the managers implement production cost reductions, greater employee effort and supervisory activities, or else the earning capacity of the corporation declines, which leads to reactions on the capital market as described above.

The competition among managers for scarce executive positions also reinforces the need for managers to conform to efficiency targets: their economic successes/failures very strongly influence their own market value on the job market. Thus, the managers are forced to operate in the interests of ownership in order to satisfy their own income-related interests.

Further arguments for reducing control costs point out that the close linkage of the manager's income to corporate profits is an incentive for efficient performance. Similarly, an appropriate organizational form (such as the multidivisional firm) can contribute to lower management evaluation costs.[16)]

The constraints on manager discretion as discussed here have convinced the property rights theorists to consider the publicly-held corporation as an efficient form of enterprise, especially since this form of enterprise is the result of voluntary agreements among

---

16) cf. in particular Williamson (1975), pp. 132 ff.

resource owners. Accordingly, the advantages of jointly raising capital, risk sharing and specialization of management functions are not offset by losses of internal efficiency. According to discussions in comparative empirical studies of owner and manager-controlled corporations, no differences in productivity or profitability can be discerned.

3. The Controversy in Light of the Results of Empirical Research

The preceding discussion has presented the two opposing views (i.e. managerialism and the property rights theory of the firm) on the relative efficiency of the publicly-held corporation. The large publicly-held corporation has not only been the subject of study for many theoretical discussions, but it has also been studied in numerous empirical investigations. The following sections will explain the basic structure of these studies and provide an overview of the results in order to provide some clarification of the contradictory assertions.

3.1 The Basic Structure of Empirical Studies

Berle and Means were working empirically as early as 1932. In this case, it was a descriptive study which determined the percentages of manager-controlled and owner-controlled corporations among the largest corporations in the U.S. Further studies of this type followed, for example, by Gordon (1945), Florence (1961) and Villarejo (1962, 1963), and, for West Germany, by Steinmann/Schreyögg/Dütthorn (1983).

The first comparative efficiency studies, carried out on U.S. corporations, were made in the late 1960's. These studies, along with most subsequent studies of this kind, are structured according to the basic study design as shown in Figure 1 (see next page).

The initial part of the studies consists of selecting one or more target variables (T), as shown in Figure 1: e.g. return on equity, total assets and/or stock, etc. For these target variables, the managerial theories have predicted lower attributes for the manager-controlled corporation compared to the owner-controlled corporation, while property rights theory has predicted no difference for either

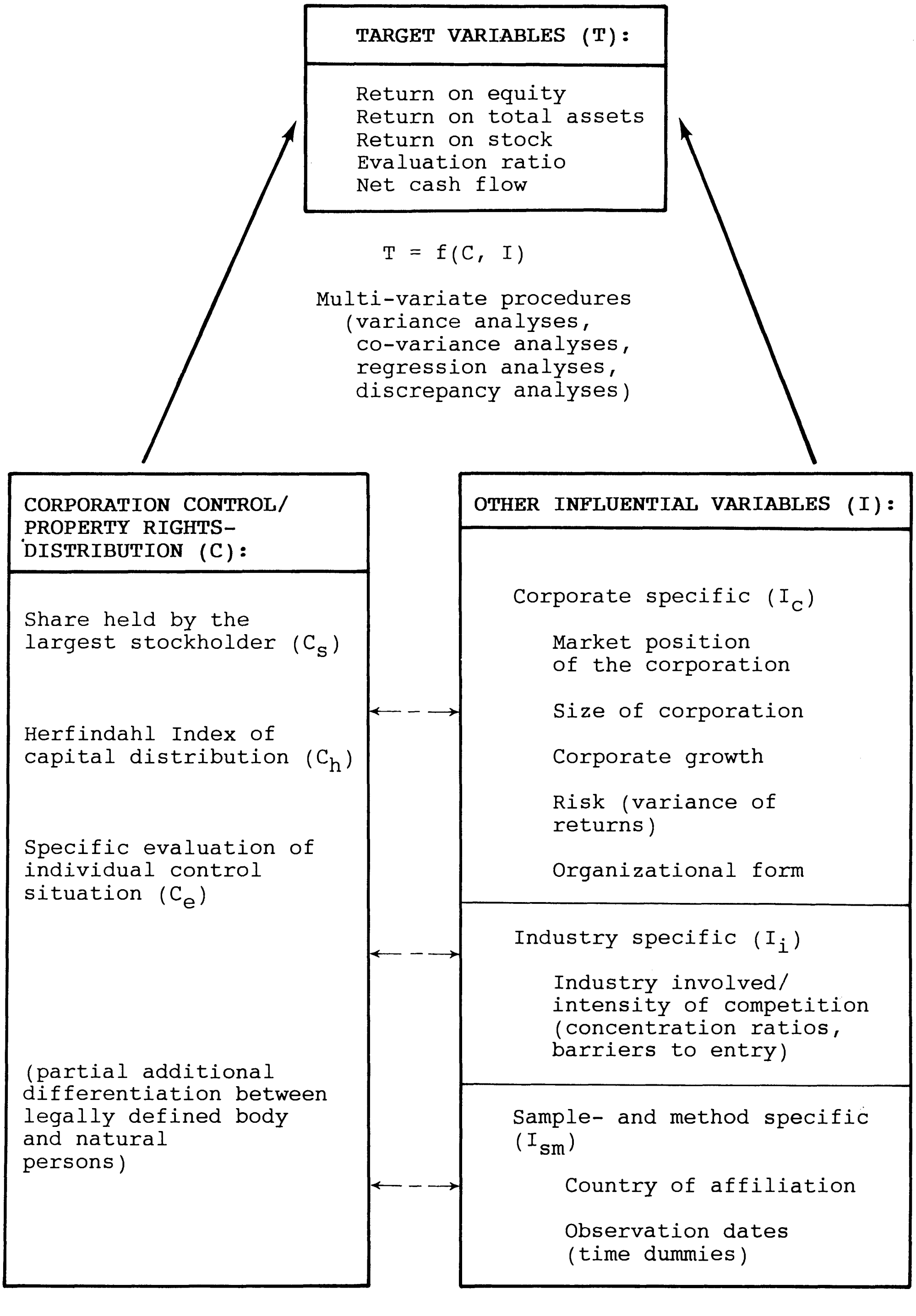


Figure 1: Schema outlining the basic research design of empirical comparative efficiency studies into the separation of ownership and control

private enterprise group. For this reason, the studies attempt to determine the statistical influence of corporate control (C) on the above-mentioned target variables (T). Several other corporate-specific or industry-specific influential variables (I) of the target variables (T) are usually taken into account in the multi-variate procedures used in the studies, in order to isolate the influence of the corporate control variables (C) on the target variables (T). This amounts to estimating the mathematical function T = f(C, I) (see Figure 1).

Several proxies have been used for corporate control (C): share held by the largest stockholder (C_s), Herfindahl Index of capital distribution (C_h) or specific evaluation of individual control situation (C_e). In most cases, corporate control (C) is operationalized with the help of the capital share of the largest stockholder(C_s). First a limit value is defined (for example 25%); if a corporation has one owner holding more than 25% of the capital, it is classified as owner-controlled (concentrated property rights), since a capital owner with more than 25% is given the possibility of effective control over management. Alternatively, corporations in which the largest stockholder holds less than 25% are classified as manager-controlled (attenuated property rights). Some studies additionally take into account whether the largest owner of capital is a natural person or family, or whether it is a corporation. If, for example, a corporation holds 40% of the capital of another corporation, the latter cannot be classified as owner-controlled without further investigation, since it is theoretically unclear whether corporations in the role of majority owner behave as owners.[17] In studies considering this problem, these cases are either excluded,[18] are considered separately[19] or are treated in a "second-stage of analysis", as proposed by Steinmann and others. In this second-stage of analysis, those corporations whose major shareholder is another corporation are classified in the control category of the parent corporation. It is assumed that the behavior of the controlling corporation is in line with its own internal control procedures.[20]

---

17) cf. for example Monsen/Chiu/Cooley (1968), p. 438; Thonet (1977), p. 153; Schreyögg/Steinmann (1981); Steinmann/Schreyögg/Dütthorn (1983); Picot/Michaelis (1984), pp. 258 ff.; Kaulmann (1987), pp. 124-128.

18) cf. for example Monsen/Chiu/Cooley (1968); Thonet (1977).

19) cf. Kaulmann (1987), pp. 162-210.

20) the second stage of analysis is strongly criticized by Picot/Michaelis (1984).

In some of the empirical studies, the corporation is not classified in control categories according to "rigid" limit values for the share of the largest stockholder (C_s). Instead, whether or not management is effectively controlled by ownership in each corporation is ascertained on the basis of interviews, newspaper reports, records of general meetings, etc. (C_e).[21] Thus, an investigation is carried out in each individual case in order to record the "true" control situation. In contrast, the "limit value-method" (C_s) directly measures the effect of distribution of capital, i.e. whether a corporation is less profitable if capital is distributed among many stockholders. In the context of property rights theory, the latter method seems appropriate, since distribution of property rights is determined by distribution of capital shares, which induces varying economic behavior. This places greater emphasis on the possibilites for exerting influence and on control through the existing distribution of rights. For this reason, it seems appropriate that almost all studies define the control situation on the basis of a largest-stockholder limit value (C_s).

Some studies do not ascertain the control situation (C) as a discrete variable (manager-controlled vs. owner-controlled), but rather determine the concentration of capital shares as a continuous variable by calculating the Herfindahl Index (C_h).[22] In principle, this procedure takes more information regarding the distribution of property rights into account compared to dichotomization: for example, a differentiation is made between corporations in which the largest stockholder has 26% or 96%, whereby with a rigid limit value (C_s)both of these corporations fall into the category of "owner-controlled." Analogously, the Herfindahl Index takes into account whether a corporation has 5 stockholders each holding 20%, or whether the stock is distributed among thousands of small stockholders. In contrast, the methods using a rigid limit value uniformly classify these corporations as "manager-controlled." The greater amount of information taken into account probably leads to varying conclusions when carrying out correlation calculations, even though in principle the pooling of corporations into two classes should also detect differences regarding returns and productivity. As a result, the use of a concentration variable in the discussed empirical studies appears

---

21) cf. Jacquemin/DeGhellinck (1980); Witte (1981a, 1981b).

22) cf. Böbel/Dirrheimer (1984); Bühner (1985).

to be desirable but not necesssary. Those empirical studies which use a continuous variable for share distribution will be considered separately in the following section.

In addition to the question whether the corporation is owner- or manager-controlled (C), almost all studies examine the effect of other potential influential variables (I) on the target variables (T). The other influential variables (I) could overlap the studied relationship between dependent variables (T), such as return or productivity, and the independent variable control category (C), with the effect that the true influence is not measured in the studies. To a great extent, selection of the variables (I) is in line with the relevant studies of industrial economics[23], which studied among other subjects the influence of different market structure elements on corporate returns. The further influential variables (I) used in the empirical studies of separation of ownership and control can be divided into three categories (see Figure 1 above):

(i) Corporate-specific (I_c): These variables are ascertained for each individual corporation. As a result, the effects on return, as theoretically explained in other studies, of the corporation`s market position (measured for example as market share), the size of the corporation, risk or organizational form can be taken into account.

(ii) Industry-specific (I_i): These variables are assigned to individual cases on the basis of the industry involved. Here, particular importance is given to the possible effects of competitive pressure to which the company will be exposed in the industry. The competition variable is frequently operationalized using concentration ratios or barriers to entry.

(iii) In addition, there are sample-specific and method-specific variables (I_{sm}): for example, time dummies are incorporated in the methods in order to record any variable coefficient level shifts in the course of the time period being studied, if a purely cross-sectional analysis is not involved.

Together with the control variables (C), these variables are taken into account in multi-variate procedures. Furthermore, covariates or

[23] cf. as overview Scherer (1980); Kaufer (1980).

interaction terms may be included in the calculations, in order to record any possible interactions between the influential variables. In part, the variables are not applied according to the figures being studied, but rather are transformed in some mathematical form. This is especially true for the corporate size variable, whose reciprocal value, logarithm, root or square can be incorporated. Some of the studies even utilize several of these variables. In this context, it should be noted that even the corporate size variable may be used as a transformation figure of the other variables to counter the heteroscedasticity problem.

Studies structured in this way in each case calculate the coefficients of all the independent variables (C, I). They then evaluate the effect of the particular control-variable (C) on the target variables (T) of the corporations on the basis of the mathematical sign, the order of magnitude and the significance of the coefficients of the control-variables (C).

3.2 An Overview of Study Results

The framework described in the previous section has provided a structure for two surveys[24] of empirical investigations into the separation of ownership and control. Figure 2 (see next page) summarizes the results of these two surveys.

In order to assess conclusively the extent to which either managerialism or the property rights theory of the firm is empirically confirmed, all the empirical studies would have to be presented in detail and analyzed to determine their respective strengths and weaknesses. This is beyond the scope of the present report. Therefore attention will be given to some specific features and general trends in the empirical literature, and conclusions reached in the analytical literature will also be examined.

A total of three out of 23 studies were classified as having reached "no clear conclusion." Child was included because his study reached different conclusions for different industries. Stano and Jacquemin/DeGhellinck were included because they each used two model specifications, each of which led to different results. In both

[24] cf. Thonet (1977), pp. 29-32; Kaulmann (1987), pp. 96-100.

Owner-controlled corporations show higher returns	Owner-controlled corporations do not show higher returns	No clear conclusion
Monsen/Chiu/Cooley (1968)	Kamerschen (1968)	Child (1973)
Larner (1970)	Elliot (1972)	Stano (1976)
Hindley (1970)	Sorenson (1974)	Jacquemin/
Radice (1971)	Holl (1975)	DeGhellinck
Palmer (1973)	Ware (1975)	(1980)
Boudreaux (1973)	Qualls (1976)	
McEachern (1975)	Kania/McKean (1976)	
	Thonet (1977)	
	Koshal/Pejovich (1978)	
	McKean/Kania (1978)	
	Witte (1981a, 1981b)	
	Böbel/Dirrheimer (1984)	
	Bühner (1984, 1985)	
	Kaulmann (1987)	

Figure 2: Results of comparative efficiency studies for the separation of ownership and control

studies, the approaches proving that owner-controlled corporations show higher returns than do manager-controlled corporations are assigned greater significance. Kaulmann shows that the arguments used by the authors to justify their model specifications are not logical.[25] On the contrary, the approaches denounced in each case by the authors, and which were also confronted with empirical evidence and showed no return differences for the corporate groups under consideration here, appear to be the better arguments. For this reason, both studies have been classified here as having reached no clear conclusion.

Among studies using a Herfindahl Index for measuring corporate control were those by Böbel/Dirrheimer (1984) and Bühner (1985). Neither study was able to ascertain higher returns for owner-controlled corporations. These findings speak strongly in favor of the property rights theory, since more information is taken into consideration when

25) cf. Kaulmann (1987), pp. 101-108.

recording the control situation with the Herfindahl Index, which in turn increases the validity of the study. In general, it appears that results are not dependent on the selected proxy for the control situation (C), since widely varying largest-stockholder limit values for classifying a corporation as owner-controlled can be found in all columns in Figure 2.[26)]

As a whole, the studies vary considerably in terms of their research design: the number of corporations included in the sample ranges from 38 to 500 and the study time periods range from 3 to 18 years. The oldest studies are based on developments in the 1950's, while the most recent studies evaluate events of the second half of the 1970's and first half of the 1980's. In addition, various countries have been investigated, and widely varying independent variables have been used in the different statistical procedures. For these reasons, it is very difficult to compare the results of the studies. Furthermore, it is very difficult (if not impossible) to determine which research design is superior in an empirical investigation into the separation of ownership and control. Hence, it is not possible to derive final conclusions regarding the empirical validity of one theory or the other. However,we are able to make some rather tentative conclusions from recent empirical evidence.

If the chronological development of the studies classified in Figure 2 is considered, the following trend becomes apparent. Until 1976 the numbers of studies for or against one theory or the other are almost identical, as has already been determined by Thonet in his study.[27)] Thonet himself finds that his study does not confirm managerialist conclusions, and thus starts a trend which has remained steady over the last ten years. All studies which have been carried out and published since 1975 show <u>no</u> differences in returns between owner-controlled and manager-controlled corporations. As a result, in the meantime the number of studies which support the property rights theory is significantly greater than the number of contradictory studies. This trend may result from, on the one hand, modification of competitive conditions, and on the other hand, from possible improvements in study design.

---

26) cf. extensive discussion in Kaulmann (1987), pp. 109 ff. in this regard.

27) cf. Thonet (1977), p. 13.

The property rights theory had based its anti-managerialist assertions primarily on the effects of competition. The corporation is subject to growing competition pressure from the following sources: increased internationalization of trade on the level of large corporations; increasing know-how on the part of countries on the threshold of industrialization; partial improvement of market transparency as a result of new information and communications technologies or institutions which lower transaction costs (e.g. consumer report journals, comparative tests in specialist journals); and acceleration of technical development. As a result of these changes, managers are not able to persistently pursue their own objectives. The uprise of the manager-controlled corporation has quite possibly prompted ownership to adopt measures in the last twenty years to reduce control costs (such as the introduction of multidivisional organizational structures and the coupling of management income to profit).

The fact that most recent results favor property rights theory may also be the result of improved study design. Perhaps some of the errors found in older studies, which led to contradictory results, have been eliminated. Whether this has actually been the case or not cannot be clarified in the context of the present study.

4. Conclusions

The most recent empirical studies into the effects of the separation of ownership and control on corporate returns indicate that the property rights theory is superior to managerial theories. This is made clear not only by its higher predictive value, which this work is able to substantiate, but also by the broader applicability opened up by the property rights theory: economic analysis of corporate charters using the property rights theory allows monitoring of different corporate forms in regard to their economic efficiency, while still maintaining the publicly-held corporation with its widely scattered stock of capital as a special area of study.

Empirical verification of the property rights theory also leads to the conclusion that the large publicly-held corporation is an efficient form of enterprise. The advantages of improved possibilities of raising capital and improved risk-sharing among investors, which are expected from this type of enterprise, are not outweighed by sacrifices of economic efficiency, which would become especially apparent through reduced returns on equity. However, this result arises only if the corporation is formed under conditions of extensive contractual discretion. This condition allows the voluntary transfer of property rights, which ceteris paribus leads to a high degree of corporate-specific economic efficiency. It is only in this way - through the voluntary exchange of property rights - that the resources of a corporation can achieve optimal utilization in an economy.

References

Alchian, A.A. (1965): Some Economics of Property Rights, Il Politico 30, 816-829

Alchian, A.A. (1984): Specifity, Specialization and Coalitions, Zeitschrift für die gesamte Staatswissenschaft 140, 34-49

Alchian, A.A.; Demsetz, H. (1972): Production, Information Costs, and Economic Organization, American Economic Review 62, 777-795

Alchian, A.A.; Demsetz, H. (1973): The Property Rights Paradigm, Journal of Economic History 33, 16-27

Baumol, W.J. (1967): Business Behavior, Value and Growth, New York

Berle, A.A.; Means, G.C. (1932): The Modern Corporation and Private Property, New York

Böbel, I.; Dirrheimer, M. (1984): Eigentumsrechte, Managementmotivation und Marktverhalten, Ein Beitrag zur Erweiterung der Theorie der Unternehmung, in: Neumann, M. (Ed.): Ansprüche, Eigentums- und Verfügungsrechte, Schriften des Vereins für Socialpolitik 140, 157-187

Boudreaux, K.J. (1973): "Managerialism" and Risk-Return Performance, Southern Economic Journal 39, 366-372

Bühner, R. (1984): Rendite-Risiko-Effekte der Trennung von Eigentum und Leitung im diversifizierten Großunternehmen, Zeitschrift für betriebswirtschaftliche Forschung 36, 812-824

Bühner, R. (1985): Internal Organization and Returns: Analysis of Large Diversified German Corporations, in: Schwalbach, J. (Ed.): Industry Structure and Performance, Berlin, 197-222

Child, J. (1973): Managerial and Organizational Factors Associated with Company Performance, Working Paper No. 4, Oct. 73, The University of Aston Management Centre, Maple House, 158 Corporation Street, Birmingham, B 6TE

De Alessi, L. (1983): Property Rights, Transaction Costs, and X-Efficiency: An Essay in Economic Theory, American Economic Review 73, 64-81

Demsetz, H. (1967): Toward a Theory of Property Rights, American Economic Review 57, 347-359

Demsetz, H. (1983): The Structure of Ownership and the Theory of the Firm, Journal of Law and Economics 26, 375-390

Elliot, J.W. (1972): Control, Size, Growth, and Financial Performance in the Firm, Journal of Financial and Quantitative Analysis, 1309-1320

Florence, P.S. (1961): Ownership and Success of Large Companies, London

Furubotn, E.G.; Pejovich, S. (1972): Property Rights and Economic Theory: A Survey of Recent Literature, Journal of Economic Literature 10, 1137-1162

Gäfgen, G. (1984): Entwicklung und Stand der Theorie der Property Rights: Eine kritische Bestandsaufnahme, in: Neumann, M. (Ed.): Ansprüche, Eigentums- und Verfügungsrechte, Schriften des Vereins für Socialpolitik 140, Berlin, 43-62

Gordon, R.P. (1945): Business Leadership in the Large Corporation, Washington D.C.

Hindley, B. (1970): Separation of Ownership and Control in the Modern Corporation, Journal of Law and Economics 13, 185-221

Holl, P. (1975): Effect of Control Type on the Performance of the Firm in the U.K., Journal of Industrial Economics 23, 257-271

Jacquemin, A.; DeGhellinck, E. (1980): Familial Control, Size and Performance in the Largest French Firms, European Economic Review 13, 81-91

Kamerschen, D.R. (1968): The Influence of Ownership and Control on Profit Rates, American Economic Review 58, 432-447

Kania, J.J.; McKean, J.R. (1976): Ownership, Control, and the Contemporary Corporation: A General Behavior Analysis, Kyklos 29, 272-291

Kaufer, E. (1980): Industrieökonomik, München

Kaulmann, T. (1987): Property Rights und Unternehmungstheorie - Stand und Weiterentwicklung der empirischen Forschung, München

Koshal, R.K.; Pejovich, S. (1978): A Note on the Separation of Ownership from Control, Management International Review 18, 41-46

Larner, R.J. (1970): Management Control and the Large Corporation, New York

Leipold, H. (1978): Theorie der Property Rights: Forschungsziele und Anwendungsbereiche, Wirtschaftswissenschaftliches Studium 7, 518-525

Marris, R. (1964): The Economic Theory of Managerial Capitalism, London

McEachern, W.A. (1975): Managerial Control and Performance, Lexington 1975

McKean, J.R.; Kania, J.J. (1978): An Industry Approach to Owner-Manager Control and Profit Performance, Journal of Business 51, 327-342

Meyer, W. (1983): Entwicklung und Bedeutung des Property Rights-Ansatzes in der Nationalökonomie, in: Schüller, A. (Ed.): Property Rights und ökonomische Theorie, München, 1-44

Monsen, R.J.; Chiu, J.S.; Cooley, D.E. (1968): The Effect of Separation of Ownership and Control on the Performance of the Large Firm, Quarterly Journal of Economics 82, 435-451

Palmer, J.P. (1973): The Profit-Performance Effects of the Separation of Ownership from Control in Large U.S. Industrial Corporations, Bell Journal of Economics and Management Sciences 4, 293-302

Picot, A. (1981): Der Beitrag der Theorie der Verfügungsrechte zur ökonomischen Analyse von Unternehmungsverfassungen, in: Bohr, K.; Drukarczyk, J.; Drumm, H.J.; Scherrer, G. (Eds.): Unternehmungsverfassung als Problem der Betriebswirtschaftslehre, Berlin, 153-197

Picot, A.; Kaulmann, T. (1985): Industrielle Großunternehmen in Staatseigentum aus verfügungsrechtlicher Sicht, Zeitschrift für betriebswirtschaftliche Forschung 37, 956-980

Picot, A.; Michaelis, E. (1984): Verteilung von Verfügungsrechten in Großunternehmungen und Unternehmungsverfassung, Zeitschrift für Betriebswirtschaft 54, 252-272

Qualls, D.P. (1976): Market Structure and Managerial Behavior, in: Masson, R.T.; Qualls, D.P. (Eds.): Essays on Industrial Organization in Honor of Joe S. Bain, Cambridge, Mass., 89-104

Radice, H.K. (1971): Control Type, Profitability and Growth in Large Firms: An Empirical Study, Economic Journal 81, 547-562

Scherer, F.M. (1980): Industrial Market Structure and Performance, 2nd edition, Chicago

Schreyögg, G.; Steinmann, H. (1981): Zur Trennung von Eigentum und Verfügungsgewalt, Eine empirische Analyse der Beteiligungsverhältnisse in deutschen Großunternehmen, Zeitschrift für Betriebswirtschaft 51, 533-557

Stano, M. (1976): Monopoly Power, Ownership Control, and Corporate Performance, Bell Journal of Economics 7, 672-679

Steinmann, H.; Schreyögg, G.; Dütthorn, C. (1983): Managerkontrolle in deutschen Großunternehmen - 1972-1979 im Vergleich, Zeitschrift für Betriebswirtschaft 53, 4-25

Sorensen, R. (1974): The Separation of Ownership and Control and Firm Performance: An Empirical Analysis, Southern Economic Journal 39, 145-148

Tietzel, M. (1981): Die Ökonomie der Property Rights: Ein Überblick, Zeitschrift für Wirtschaftspolitik 30, 207-243

Thonet, P.J. (1977): Managerialismus und Unternehmenserfolg. Ein empirischer Beitrag, Saarbrücken

Villarejo, D. (1962): Stock Ownership and the Control of Corporations, Part I and II, New University Thought 2, 33-77

Villarejo, D. (1963): Stock Ownership and the Control of Corporations, Part III, New University Thought 3, 47-65

Ware, R.F. (1975): Performance of Manager- Versus Owner-Controlled Firms in the Food and Beverage Industry, Quarterly Review of Economics and Business 15, 81-92

Williamson, O.E. (1964): The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm, Englewood Cliffs, N.J.

Williamson, O.E. (1975): Markets and Hierarchies: Analysis and Antitrust Implications - A Study in the Economics of Internal Organization, London and New York

Williamson, O.E. (1985): The Economics of Capitalism - Firms, Markets, Relational Contracting, New York, etc.

Witte, E. (1981a): Der Einfluß der Anteilseigner auf die Unternehmenspolitik, Zeitschrift für Betriebswirtschaft 51, 733-779

Witte, E. (1981b): Die Unabhängigkeit des Vorstandes im Einflußsystem der Unternehmung, Zeitschrift für betriebswirtschaftliche Forschung 33, 273-296

Contract, Agency, and the Delegation of Decision Making

Erich Schanze

Summary: Contract and Agency are viewed as separable patterns for structuring transactions in the process of the division of labor. Whereas 'contract' refers to a concept of discrete exchange with unrevealed individual choices, 'agency' relates to a concept of delegation of choice, and hence an explicit treatment of the rules of choice and preference formation. In this view agency is not a special case of a theory of contract incentives with interchangeable partners, but a concept of incentives to align the agent's preference sets and future choices with that of a principal. In consequence, behavioral aspects like trust, loyalty, opportunism (which are irrelevant in the pure pattern of contract) are central in the concept of agency. The patterns of contract and agency overlap to a large degree in the real world. An agency relationship may be based on a contract, and a contract may contain agency features ("relational", "expanded", "idiosyncratic" contracting). For purposes of analysis, however, the properties and boundaries of the two patterns should be observed.

Table of Contents:

1. Introduction: Equivocations in the History of Agency Theory

The short history of agency theory is a history of equivocations. In his paper on "The Nature of the Firm" Coase (1937) had proposed to

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

view market (contract) and hierarchy (firm) as basic organizational alternatives. These alternatives, he asserts, are selected by economic actors in a process of substitution at the margin which is guided by transaction cost considerations. In contrast, Alchian and Demsetz (1972) insisted that the firm should be regarded as nothing more than a "contractual form". When the contractual "nexus" explanation of the firm attained prominence through its elaboration by Jensen, Meckling (1976), Fama (1980) and Fama and Jensen (1983), it was widely held that Coase's original statement was "tautological" (cf. Williamson 1981). It is, however, noteworthy that the focus of theoretical attention did not remain within the original realm of "contract". It shifted to the strange concept of "agency".

Why do economists talk about "agency" if the firm is just a special cluster of contracts between some owners of resources used in a joint effort? Why did they part from the original contract/property rights framework supplied by Alchian and Demsetz? Is agency the new tautology, a mere deviation of contract theory? Are we already travelling on a return ticket from Agency to Contract? In some recent applications of agency theory it appears to be a mere matter of labelling and convenience whether the term contract or agency is used. The parties to the contract are named principal and agent, and they are, for all relevant purposes, interchangeable. However: pluralitas non est ponenda sine necessitate. Ockham's law of intellectual (and verbal) parsimony reminds us that - as a matter of logic - we should not use different terms for the same problem. Bertrand Russell (1956, 326) transfers Ockham's "rasor" into a positive rule for rational reconstruction of reality: "Wherever possible, substitute constructions out of known entities for inferences to unknown entities." Why do we worry about agency if the concept of contract - familiar to economists since Adam Smith - covers the case?

The purpose of this paper is to stimulate reflection about the substance of a possible division of contract and agency. What perceptions of reality do we entertain if we talk about contract or agency? What are the formal aspects of these patterns? Do they contain separable properties if employed in the process of planning? How can we generalize their legal structure without being captured either by received legal doctrine or by its recent ambitious economic counterpart? What are the necessary rules of maintaining a contractual or agency relationship? Do individual preferences matter? What is the basis of legitimacy for each of the two relationships? I shall give a tentative answer to these questions in a matrix at the end of this paper which is thought to serve as a starting point for further discussion.

2. Agency - A Special Case of Contract?

Let us start with the conventional view that agency is simply a special case of contract. A civil lawyer may be reminded of a proverb used some centuries ago when the catalogue of enforceable contracts was closed. If a contract could not be subsumed under one of the regular types, the practitioner relied on a rule of thumb which ran as follows: "What you cannot explain, simply call a mandate (agency)." A more reasonable approach might be to look at the specific features of the "contract" called agency. Jensen and Meckling (1976) seem to follow this "special case" theory if they define "an agency relationship as a contract under which one or more persons (the principal[s]) engage another person (the agent) to perform some service on their behalf which involves delegating some decision making authority to the agent." The second part of the definition that agency refers to the delegation of decision making authority on behalf of the agent is generally accepted. How about the contractual "foundation"?

The special case approach starts from the assumption that mutual assent is a precondition of the concept of agency. There is a meeting of the mind between agent and principal in the sense that agent will carry out a specific part of principal's business in the interest of principal. The focus is on the consensual terms of mandating. A number of cases might be conveniently explained within this figurative concept. Some language in legal decisions - from which the common law concept of agency originally derives - would support this consensual notion of agency. In Tarver, Steele & Co. v. Pendleton Gin Co. (Tex. Civ. App. 25 S.W. 2nd 156, 159) the court describes agency as the consensual relation existing between two persons by virtue of which one is subject to the other's control. It is, however, largely undisputed in the legal world that agency may be created by contract or by law. "It is not essential to the existence of authority that there be a contract between the principal or agent or that the agent promises or otherwise undertakes to act as agent" (§ 26 note [a] Restatement of Agency [1958]). Consider the case of guardianship where there is clearly no "contract" between guardian and infant or other ward. This paper does not purport substituting legal definitions for economic definitions. Specifically, it does not want to repeat the charge, recently made by Robert Clark (1985), that the "classical" relation on which modern economic agency theory is based - the relation between stockholder and manager - does not constitute an agency in a proper legal meaning. From a legal point of view this is correct. However, every social science is free to define its own terms if they are useful for explanatory purposes. Nevertheless, an economist may be well advised looking at the established meaning of a term in the other discipline in order to acquire a clearer conception of the practial implications and the context in which the term is used to describe a specific institution.

Indeed, one may reconstruct every relation imposed by law as a "contract" in the sense of a voluntary exchange transaction. The question remains whether this extension of contract to a notion of "social contract" (which has a long tradition in the political sciences) is of much explanatory value in reality. We frequently find "contractarian" explanations less than plausible. Can a citizen of the United States claim, in a proper and meaningful sense, that he is the "principal" of the President of the United States? Of course, in election campaigns the chief executive would assure his followers that he is their "agent". We should put this to the rubric of political rhetoric. Pratt and Zeckhauser (1985), in a recent overview on principal and agent theory, tell us that in writing their paper they conceive their task as that of agents for the potential reader. Is this mere academic rhetoric? Of course, we would hope that some discipline of the "reader-principals" will be exercised within the academic discourse. But, unfortunately, the authors have not asked us in advance, and substantive "monitoring" is in vain, since the book was already published when we could start to monitor their effort. On the other hand, Pratt and Zeckhauser are beyond the suspicion of a mere academic joke because they do not share the original "contract" conception of agency. They define broadly: "Whenever one individual depends on the action of another, an agency relationship arises. The individual taking the action is called the agent. The affected party is the principal." If there is more behind this definition than the truism that we are free to conceive the state of the world in more or less attenuated agency relations (a definitional variant of a common social science model of interdependence), we have to take a second look.

3. Delegation and Discretion as Central Elements of Agency

It may not be by accident that Coase distinguished the concept of "the firm" from the concept of contract; likewise, that Jensen and Meckling adopted the legal term "agency" in this context although they insisted in a "contractual" explanation of the firm. The conventional argument for this shift of concepts is summarized by Kronman and Posner (1979, 39): "The substitution of the firm for contract is the substitution of employment contracts whereby the entrepreneur pays the worker for the right to direct his work, for contracts specifying price and output by leaving the details of the work to the worker, unsupervised by the entrepreneur who is purchasing his output."
This explanation is superficial. It stresses the contractual nature of the employment relation (which shall not be denied) instead of pointing out that the gist of this relation lies in the notions of delegation and discretion. Both features are central to the concept of agency. This is precisely the link to Coase's concept of the firm, the point of "direction" or doing something by fiat in a hierarchy. Coase's conception is incomplete because it stresses direction and neglects discretion. As to the contractual nature of agency, it is true that delegation and discretion may be constituted in a contract, e.g. in an employment contract, but modern contract theory would point to the specific nature of this contract by calling it "relational" (Macneil, 1974) or "idiosyncratic" (Williamson, 1979). Goldberg (1976) correctly emphasizes that such a concept of contract is "expanded" beyond the classical notion. Indeed, the "expanded" character of these contracts is mainly vested in the presence of agency elements and the institutional problems associated with agency which are alien to the classical concept of contract.

4. The Concept of Contract

The familiar concept of contract refers to the exchange transaction normally illustrated by buyer and seller exchanging goods against money; the example may be extended to rendering services against money if the services and the mode of payment are adequately specified. With increasing complexity of the object of the contract and the extended duration between the mutual acts of performance, legal aspects of drafting (specifying) and enforcing the contract become relevant. In the model we still perceive isolated, antagonistic traders who make choices which are thought to reflect their individual preferences. These individual preferences are deemed as given; they are no explicit theme of the contract; rather, they remain implicit. Parties are indifferent towards each other's preferences and future choices. It suffices that the contracting parties agree on the contract price and that they signal their willingness to perform as specified (to pay, to deliver, to clean the window). The implicitness of preferences is also the basis of the economic perception of value. The contract price presents a veil for personal preferences and neutralizes the commitment of the parties. Lawyers cherish the general principle that the "motives" of offer and acceptance are irrelevant.

Moreover - in the model - a contract does not require a set of legal rules including sanctions for deviant behavior. A fully specified contract is self enforcing. Pacta sunt servanda is no necessary but an expedient rule.

5. The Concept of Agency

In contrast, personal choice and preference formation are central in the concept of agency. The core of agency, it is submitted, is the delegation of personal choice to somebody else; preferences and future choices of the agent are the explicit theme of agency. Instead of indifference towards the other party's preferences, concern for the other's interest, his personal commitment and loyalty vis à vis my perceived or real choices, become an operative feature of the relationship.

Agency is delegation of choice, or rendering discretion. There is a necessary ongoing interdependence of the utility functions of principal and agent. Whereas contract is performance oriented, agency is effort oriented.

There may be a complete delegation which is present in the typical non-contractual variant of agency, that of guardian and ward. The ward has no relevant personal preferences. Extensive forms of delegation and discretion and, hence, concern are typically found where the agent offers highly specialized and valuable human capital to be invested in principal's affairs. Trust and confidence in the competence

of the agent replace detailing performance. Professionals like doctors and lawyers tend to accept mandates to act without negotiating about the details of their undertaking.
Delegation and discretion may be limited. The workers know what to do in Adam Smith's pin factory. However, they may have strong preferences for leisure, hence, shirking and monitoring. The relationship contains agency features.

On the other hand, pure delegation without discretion is no case of agency. Every legal order distinguishes between agent and simple messenger. The messenger or the person who works strictly under the direction of the employer do not pose agency problems; they are within the realm of classical contract law. In other words: if the agent may not articulate his preferences and act according to his own choices, he is no agent but a simple contractor. In this case, it is proper to speak of a contract concerning payment against the right to direct. Employment contracts are rarely of this kind. Even if the performance is strictly defined, an agency relation may arise because the modality of payment is not strictly correlated with performance. If I pay by hour, time may be wasted. If I pay by unit, a specified quality standard may not be reached. However, even "complex" contracts such as procurement contracts may be "classical" contracts, and consequently involve no principal-agent relation. This is the case if a fixed price is specified for procuring a specific item. Once schemes for sharing risks and benefits are conceived, parties have to be concerned about the other's future choices. In a "classical" contract situation I do not care whether my counterpart has a strong preference for leisure. I will pay the student if he has cleaned my window. I do care, however, once the student enters my law firm. And he will (and should) care, too. In this bilateral (double) agency relation I am concerned whether my counterpart wastes my assets, and he is concerned whether I waste his. We may agree on a monitor, and thereby create a new agency relation. The monitor directs us and measures our productivity but he may also shirk. We may also agree on production targets, or other efficiency criteria which we monitor ourselves.

Contracts with agency features typically have high front/end costs. In other terms, agency is dependent on legal rules concerning sanctions for breach of loyalty of the agent and sanctions for non-performance of promised incentives of the principal. Expenditure is made not only to establish the capacity but also the reliability of the partner. Personal sanctions for deviant behavior are specified including dismissal, loss of bonded items, loss of reputation, etc. Problems associated with information about reliability and with specifying rights and duties of principal and agent relating to the delicate task of preference formation for a third party, require substantial institutional consideration. Compliance with preferences (that the agent will choose in my interest) is signalled by prior action, commitment, or specific contractual obligation of the agent. Behavioral aspects like trust and confidence, loyalty, absence of opportunism, which do not play a role in a classical contract environment, have to be included in institutional considerations.

The variance in modes of specifying agency relations is manifold. There may be fully specified self enforcing relations. There may be written rules of behavior in a contract or in a statute; there may be implicit rules covered in general principles such as the requirement of a fiduciary duty - typically the duty of care exercised by a prudent professional man acting in the given situation. The degree of dependence plays a role. My academic future may not be dependent on what Pratt and Zeckhauser think about agency theory - hence, no agency relation.

6. Contractual and Non-contractual Agency Relations

If the thesis is accepted that some contracts contain agency features but that agency is not a special contract, it follows that agency relations do not necessarily require a contractual foundation. Of course, a contract is a solid base of legitimizing the agency relation. Contracts involving agency elements contain their own source of legitimacy by referring to the mutual consent of the parties. An agency relation may be based on an explicit contract of delegation (mandate). However, it may also be based on a statute. The lunatic is not asked if guardianship is devised; neither may the child under normal circumstances challenge the legal representation of its parents.

Beyond contract and statute intermediate forms of legitimizing agency relations are of special interest. Various forms of voting such as majority vote, assent by presence, or dissent by exit might be a sufficient basis for legitimizing that a third person substitutes (parts of) his preferences for (parts of) my preferences and chooses on my behalf. I may not be able to challenge the board of Volkswagen but I may sell my shares. Voting is sometimes explicitly organized by general law, sometimes by corporate charter and by-laws, sometimes by contract (cf. Schanze 1986 on corporate voting). In some cases I may structure the conduct of my agent by devising or accepting a program of conduct, in other cases I may have the right of ratifying the action of a third party thereby declaring that he acted as my agent. It may be of interest to economists working in the framework of agency theory that lawyers are less concerned with the specification of incentives in agency relations (which seems to be the current interest of economists) but with the definition of the boundaries of the agency relation itself. Lawyers have developed an extensive arsenal of rules concerning the question under what conditions transactions of persons who are supposed to act on another person's behalf are regarded as having transacted in a binding fashion for that other person. In the field concerning management's representation of corporations, this problem of "identifying" the correct partner is particularly relevant. May I, the owner of a one-man corporation which is on the verge of bankruptcy, argue that buying an expensive machine ("on behalf of the corporation") was not my personal business but that of the corporation? Why may I claim that I am not the principal of this transaction but merely an agent of the corporation?

7. Conclusion: A Tentative Matrix of Contract and Agency Features

Agency theory is no special theory of contract incentives but refers to a model of delegation of preferences and the monitoring of discretion of a third party. A number of important agency problems are discussed in the framework of an "expanded" theory of contract (relational or idiosyncratic contracting). This should not blur the borderline between the original patterns of contract and agency. They are the basic variants of organizing the division of labor. Naturally, they have a large sectional area in common if one looks at real world transactions.

CONTRACT AND AGENCY

BASIC TRANSACTIONAL PATTERNS

DIVISION OF LABOR	CONTRACT	AGENCY
formal aspect	market	hierarchy
perception of actors	isolated	integrated
procedural aspect of planning	self-determination specification discreteness performance orientation	delegation representation discretion interdependence effort orientation
legal structure	mutual consent on two individual decisions to perform	unilateral authorization to act on behalf of principal
necessary rules	none (fully specified contract is self enforcing - pacta sunt servanda is no necessary but an expedient rule)	sanctions for breach of loyalty sanctions for non-performance of promised incentives
revelation of individual preferences	does not matter	central feature
real world paradigm(s)	exchange of goods and services	manager sales agent guardian sharing arrangements
basis of legitimacy	reciprocity (mutual consent of parties)	power (delegation by (1) consent (2) voting (3) statute)
entitlement structure	exchange of property rights	horizontal split-up of property right; paradigm: decision control + residual ownership / decision management (Fama, Jensen 1983)

References

Alchian, A.A.; Demsetz, H. (1972): Production, Information Costs, and Economic Organization, American Economic Review 62, 777.

Clark, R.C. (1985): Agency Costs Versus Fiduciary Duties, in Pratt, J.W.; Zeckhauser, R.J. (eds.): Principals and Agents, The Structure of Business, Harvard Business School Press, Boston (Mass.), 55.

Coase, R.H. (1937): The Nature of the Firm, Economica [n.s.] 4, 386.

Fama, E.F. (1980): Agency Problems and the Theory of the Firm, Journal of Political Economy 88, 288.

Fama, E.F.; Jensen, M.C. (1983): Agency Problems and Residual Claims, Journal of Law and Economics 26, 327.

Goldberg, V.P. (1976): Toward an Expanded Theory of Contract, Journal of Economic Issues 10, 45.

Jensen, M.C.; Meckling, W.H. (1976): Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, Journal of Financial Economics 3, 306.

Kronman, A.T.; Posner, R.A. (1979): The Economics of Contract Law, Little, Brown & Co. Boston, Toronto.

Macneil, I.R. (1974): The Many Futures of Contracts, Southern California Law Review 47, 691.

Pratt, J.W.; Zeckhauser, R.J. (1985): Principals and Agents: An Overview in Pratt, J.W.; Zeckhauser, R.J. (eds.): Principals and Agents, The Structure of Business, Harvard Business School Press, Boston (Mass.), 1.

Russell, B. (1924): Logical Atomism in Russell, B.: Logic and Knowledge. Essays (R.C. Marsh, ed.), London, New York 1956, 326.

Schanze, E. (1986): Potential and Limits of Economic Analysis: The Constitution of the Firm, in Daintith, T.; Teubner, G. (eds.): Contract and Organisation, Legal Analysis in the Light of Economic and Social Theory, Walter de Gruyter, Berlin, New York, 204.

Williamson, O.E. (1979): Transaction Cost Economics: The Governance of Contractual Relations, Journal of Law and Economics 22, 233.

Williamson, O.E. (1981): On the Nature of the Firm: Some Recent Developments, Zeitschrift für die gesamte Staatswissenschaft (JITE) 137, 675.

A Note on Fair Equality of Rules

Joachim Voeller

Summary: Fair equality of rules implies "treating equals equally and unequals unequally". The problem with this definition is the determination of criteria by which groups of equals are formed. The paper considers the constitutional relevance of this issue by considering the German Supreme Court's approach.

Table of Contents:

1. Introduction

It is a focal concern of the theory of property rights to investigate the influence of legal and institutional arrangements or regulations on the economic action of man. Since the exchange of goods and services is generally performed through contracts, the realization thereof imposes significant transaction costs. These transaction costs should be interpreted in a wide sense thus including for instance all costs inferred from the control of the contract provisions. Hence, a primary goal centres on the question on how transaction costs depend on the institutional und judicial framework and on what interrelations between those two factors exist. Therefore it is obvious that the design, for example, of incentive schemes between a principal and his agent or the fairness of the terms of a contract deserve utmost attention. Fair rules between both parties are necessary in order to reach and sustain any agreement on possible compensation payments. After all any

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

remuneration of the agent by his principal constitutes nothing but a compensation payment for certain services the agent delivered. But who defines what a fair compensation looks like?

In the following no attempt is made to present at least parts of the most important concepts of fairness or justice that have been developed over much of human history either by philosophers or different religions or in certain disciplines like jurisprudence or even economics. Rather we restrict ourselves to the German Constitution of 1949 and will show that the notion of 'fair equality of rules' is already rooted as a basic dilemma in constitutional law. But since all legal regulations must, in the last instance, satisfy the norms of the constitution, any knowledge of what criteria the German Supreme Court uses for interpreting Art. 3 (1) of the constitution (see chapter 3) may prove quite illuminating and useful for answering our question on the ingredients of fair rules. It will be interesting to see how the highest court tries to solve the antimony between just a formal and a more substantive interpretation of this basic constitutional right. Then it might be feasible to apply the same principles and ideas as juridical guidelines in adequately structured principal-agents problems. At any rate it would enable all contracting parties to discuss and possibly agree on a set of essentials to a fair contract.

2. The Problem

The fundamental difficulties in finding and formulating general criteria of fairness may be demonstrated most easily by looking at a few examples:

a) There is no doubt that an equal rule for unequal persons subject to this rule can but must not be fair. For instance the same examination for unequal candidates - they may, e.g., differ in age - can be extremely unfair but it does not have to be that way.

b) The French poet Anatole France is often cited with the following, very famous phrase: "The majestic equality of law prohibits the rich and the poor from sleeping under bridges, begging in the streets and from stealing bread". Here, legal equality means that all members of a society are subject to the same general rules (law). General rules define situations for which they hold no matter who gets into the respective circumstances. But, generality and equality of a rule apparently does not ensure the same effects on the persons it affects. Equality of rules, i.e., general and equal rules, does not guarantee the same legal or binding effects of the respective rule.

c) If all participants of a game (like chess, roulette or an athletic match) agree on "fair" rules, then the players will not apply different standards to the outcome of the game as long as the rules of the game have not been violated. In other words: If the rules of the game are considered fair, then the same holds for the results even though the effects of the rules may be extremely unequal. Hence, equal (legal) effects of a rule are not a necessary condition for fair equality of rules.

d) As we shall see the picture looks different in the next example. In their extensive paper on "Altersversorgung im Umbruch" Spremann/Zink [1986, p.35] give the following criteria for judging the merits of a pension system in a society: Political stability, intergenerational fairness, interfunctional fairness, institutional fairness .... "Intergenerational fairness (= fair equality of rules) requires that one generation must not be treated better or worse than another ....". Obviously the rule is considered to be fair if its legal effects treat all retired persons in such a way that their individual situations are adequately taken into account. Below we will explain what that means.

Now, what conclusions with respect to our notion of fair equality of rules can we draw from the observations above? First of all we note that a purely 'formal' interpretation of equality of rules meaning 'equal and general rules for all' evidently is not sufficient. In many cases such a definition will not yield fair outcomes. The reason for this drawback lies in the fact that unequals are treated equally and the consequences of such action are often very discriminating. If, however, as in example d) the same legal (binding) effects are desired then equals must be treated equally and unequals unequally. As a result different binding effects of a rule can be prevented or at least be compensated for. Such a conception of fairness would naturally exclude treating either equals unequally or unequals equally.

Finally, in the third example cases are noted where unequal binding effects are accepted because the rules under which the often quite uneven results are reached, are considered intrinsically fair ("procedure fairness"). Again this attitude crucially depends on the notion of equal treatment or chances for each, strictly speaking, unequal player.

Now, a very difficult but basic problem arises: When are equals treated equally and unequals unequally? Or to put it differently, what criteria exist to distinguisch first of all equals from unequals and, secondly, by what standards are equal binding (legal) effects of a rule to be judged?

It is obvious that any answer to both questions requires value judgements on the respective definition of equality. Since value judgements can never be proven true or false the concepts of equality with respect to both cases will greatly differ among people. For instance, several principles of fairness are discussed in public finance to tax different incomes "fairly" by burdening each taxpayer with an (relatively or absolutely) equal sacrifice in utility. Also, different groups of 'equals' are common in athletic competitions. Nobody would require men to compete with women, for example, or healthy athletes with handicapped. As indicated before all distinctions or classifications lead to the well-known phenomenon that fair equality of rules involves a formal aspect as to general and equal rules and a substantive (material) aspect as to the legal or binding effects of a rule. It is exactly this fundamental dilemma which can also be found in the German constitution and which must be dealt with by the Supreme Court justices.

3. Fair Equality of Rules as a Constitutional Dilemma

Art. 3 (1) of the German constitution states: "Alle Menschen sind vor dem Gesetz gleich" (All men are equal before the law).

The constitution warrants legal equality both in a formal and in a substantive sense in the course of which the so-called "Sozialstaatsprinzip" ("social welfare principle") of Art. 20 and 28 is often quoted to supplement a substantive interpretation of Art. 3. The lawmaker is seen to be obliged to treat equal things equally and, accordingly, unequal things unequally (see v.Mangoldt/Klein, 1957, p. 198).

The formal aspect of Art. 3 requires the validity and the enforcement of the law without exception and discrimination. "Jeder wird in gleicher Weise durch die Normierungen des Rechts verpflichtet und berechtigt, und umgekehrt ist es allen staatlichen Stellen verwehrt, bestehendes Recht zugunsten oder zulasten einzelner Personen nicht anzuwenden" (Hesse, 1977, p. 176). Still, formal equality is not interpreted as absolute equality but rather taken as relative equality.

Hence differentiations have to occur whenever the disregard of differences by the legislation can not be justified by nature and, therefore, unequal treatment would be arbitrary ("Willkürverbot", "prohibition of arbitrariness") (see Maunz / Dürig / Herzog, GG, Art. 3 (1), Rdn. 266).

Different views are expressed on how deeply substantive (material) equality is already embodied in Art. 3 (1) of the constitution. Anyhow it will unfold its effects in conjunction with the balancing and compensating power of the "Sozialstaatsprinzip". The legislation, at any rate, is obliged to establish equal law for all that is equal. As mentioned above, the lawmaker must not create artificial inequalities. Rather he has to take into consideration existing inequalities primarily those given by nature or developed in social life (see v.Mangoldt/Klein, 1957, p. 199). Hence, substantive equality of rules is given in as much as social and economic conditions preventing "fair" solutions are well considered. For instance, the lawmaker is urged to actively limit any economic power in the marketplace in order to enable the economically powerless to take advantage of liberty rights to the same extent powerful organizations do. Since natural and social endowments are quite distinct, formal equality of rules alone would provoke inequalities that could make the same use of equal rights an unrealistic vision. Therefore it must be a prime objective of government to uncover social inequalities that deserve attention since they inadequately inhibit the realization of rights formally bestowed on everybody equally. Only then fair solutions are possible.

4. The Position of the Supreme Court

In the following the basic attitude of the highest German court ("Bundesverfassungsgericht") with respect to the complex relationship between formal and material equality is briefly discussed. However, the legal instrument to enforce Art. 3, i.e. the so-called 'constitutional norm control' ("Normenkontrolle") is not commented on in this connection. The problem whose solution we are now looking for can be defined as follows: Do there exist any criteria the Supreme Court systematically applies for enforcing and controlling the equality of legal rules both in a formal and in a substantive sense? In other words, are there any top-level decision rules that can be used to further fairness through fair equality of rules? In case such standards exist, they would be of great value for example to any economic policy- and lawmaker who is often confronted with the antimony between the formal and substantive meaning of certain rules or policy measures.

With regard to Art. 3 (1) and the constitutional legitimacy of any rule the opinion is generally accepted that it is not the high court's responsibility to examine whether a law provides the most adequate, the most reasonable or even the fairest solution to a problem (see BVerfGE 9, p. 206; 14, p. 238; 17, p. 330). The Supreme Court just has to check whether arbitrary regulations have been introduced that are not justified by reason and intrinsic arguments (see Badura, 1967, p. 399 and Zacher, 1968, p. 352). "Welche Elemente der zu ordnenden Lebensverhältnisse maßgebend dafür sind, sie im Recht gleich oder ungleich zu behandeln, entscheidet grundsätzlich der Gesetzgeber" (BVerfGE 3, p. 240).

Thus, equality or inequality of rules is not something a priori given but rather the result of a normative legal process. For its legal decisions the legislative body must be endowed with broad discretion and freedom. The constitutional control then is reduced to the control of and search for arbitrary elements whose irrelevance to the regulation is obvious. This judicial self-restraint of the Supreme Court is very important and must be kept in mind when the consequences of state intervention into the free play of market forces is analysed. Very often these interventions seem to treat unequals equally or equal facts unequally. As long as no arbitrary distinctions or regulations are written into the rules the lawmaker is free to shape the respective world.

"Wer also hofft, durch die Lektüre der verfassungsgerichtlichen Entscheidungen zur Gleichheitskontrolle Aufschlüsse über die Sachstrukturen, die zur Beurteilung standen, zu erhalten,.wird im Regelfall enttäuscht" (Zacher, 1968, p. 352). In this connection the Supreme Court holds that it is not possible to state abstractly and generally what is arbitrary or irrelevant in a given situation. "Solange die Regelung sich auf eine der Lebenserfahrung nicht geradezu widersprechende Würdigung der jeweiligen Lebensverhältnisse stützt, ... kann sie von der Verfassung her nicht beanstandet werden" (BVerfGE 17, 216).

As a result the attempt to derive general criteria for fair equality of rules from the Supreme Court's interpretation of Art. 3 (1) of the constitution must be considered partially unsuccessful. Besides the prohibition of arbitrariness which itself is quite a vague concept there does not seem to exist any general standard that would release the principal, i.e. the lawmaker, from finding a sensible definition of fairness before enacting a rule.

References

Badura, P.: Die Rechtsprechung des Bundesverfassungsgerichts zu den verfassungsrechtlichen Grenzen wirtschaftspolitischer Gesetzgebung im sozialen Rechtsstaat, in: Archiv des öffentlichen Rechts, Bd. 92, 1967, p. 382 ff.

BVerfG: Von den Mitgliedern des Bundesverfassungsgericht (BVerfG) herausgegebene Sammlung der Entscheidungen des BVerfG, Bände 3, 9, 14, 17, Tübingen, div. Jg.

Hesse, K.: Grundzüge des Verfassungsrechts der Bundesrepublik Deutschland, 10. Aufl., Heidelberg/Karlsruhe 1977.

v.Mangoldt/Klein: Das Bonner Grundgesetz, Kommentar, München 1957.

Maunz, Th.; Dürig, G.; Herzog, R.: Grundgesetz, Kommentar, München o.J.

Schmiedeberg, A. u. Voeller, J.: Formale und inhaltliche Gleichheit - Zur Antinomie verfassungsrechtlicher Grundprinzipien, Discussion Paper des Instituts für Wirtschaftstheorie und Operations Research der Universität Karlsruhe, Nr. 273, Karlsruhe 1986.

Spremann, K. u. Zink, A.: Altersversorgung im Umbruch, Universität Ulm, Ulm 1986.

Zacher, H.F.: Soziale Gleichheit, in: Archiv des öffentlichen Rechts, Bd. 93, 1968, p. 341 ff.

Section 8 Agency Costs

Agency Costs and Transaction Costs: Flops in the Principal-Agent-Theory of Financial Markets

Dieter Schneider

Summary: Agency costs and transaction costs are generally used to explain agency-problems. But this means an inherent contradiction in a world of uncertainty if costs are defined as a quantitative concept. To avoid this contradiction it is suggested to ascribe only a metaphorical sense to the term "cost". However, a basic concept used as a metaphor does only verify the incompetence of scientists to get hold of their problems with the tools at hand: If cost is a quantitative concept and allocative efficiency is wanted there is no principal-agent-problem whenever agency costs and transaction costs can be calculated and there are no costs whenever an agency-problem exists.

Table of Contents:

1. The Agency-Problem and the Ways of its Solution

The term "principal-agent-problem" suggests a wrong idea about the facts to explain: The actual concern is not the relationship "master (principal) and servant (agent)". For as long as knowledge is power, the superior power is with the agent: "Princip der Vorhand" said Wilhelm Roscher (21, p. 362 f.) 13o years ago.

The principal-agent-problem is a central issue in a theory of organizing human relations: There is always a division of labour within the human society. Only hermits are completely self-sufficient people and not capable of reproduction. As soon as someone does not eke out his existence exclusively for himself and is living together

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

with other people his fellow men allocate responsibilities to him and provide him with rights, i.e. claims for the realization of tasks by others.

When men assign tasks to each other and take over rights and duties, acting on someone's behalf and instructions comes into existence. This is fashionably named the principal-agent-problem:

An instructor, called principal, hires another person, the agent, for the realization of a special task. In simple cases there is only a problem of risk-sharing between persons with the same information (identical probability beliefs). But in general the principal cannot observe the action chosen by the agent and therefore the information between principal and agent is asymmetric. The agent has a superior knowledge or gains it on his job. Varus who led the Roman Legions of Augustus into the marshy Westphalian woods already had the superior knowledge compared with the Roman Emperor far away. And his principal Augustus could not watch the scope of action he had and the chances he missed against Herman the leader of Cherusci, one of the Teuton tribes, who defeated Varus.

One approach to handle the economics of the principle-agent relationship, especially with regard to its paradigm of separation of ownership and control, is the concept of agency costs as proposed by Jensen/Meckling (14) and Fama/Jensen (1o). An analogous approach is the concept of transaction costs (27).

Both approaches seem to be more familiar to economists and more related to applications than the difficult mathematical analysis to some kind of principal-agent relationship by Radner (19), Grossman/Hart (12) or in the survey by Rees (2o).

But the familiar opportunity cost approaches to institutional arrangements, explicated in the concepts of agency costs or transaction costs, seem to be flops, if the three theses of this paper are valid:

1. Whenever agency costs in the sense of Jensen/Meckling/Fama can be calculated there is no need for monitoring, whenever there is a need for monitoring, agency costs cannot be calculated.

2. Neither the concept of transaction costs nor any other opportunity cost approach can explain (in the sense of yielding testable hypotheses), why institutional arrangements have to be constructed, because opportunity cost as a quantitative term and observable fact implies a market system in equilibrium. But equilibrium theory defines away institutions like the principal-agent relationship.

3. If entrepreneurial functions are required in arbitraging processes in markets or to realize economic development in a dynamic Schumpeterian sense (24, p. 481-483), then institutional arrangements like the principal-agent-relationship should be explained by taking over entrepreneurial functions in market processes, not by equilibrium theory and its offspring of opportunity cost approaches.

2. The Flop Named "Agency Costs"

Agency costs are defined as the sum of (14, p. 3o8):

(1) the monitoring expenditure by the principal,
(2) the bonding expenditures by the agent,
(3) the residual loss, i.e. the monetary equivalent of the reduction in welfare experienced by the principal due to the divergence between the agent's decisions and those decisions which would maximize the welfare of the principal.

The first two kinds of costs are a consequence of the effort to minimize the residual loss. Therefore, the definition of agency costs puts the cart before the horse. Properly, the idea should be expressed in the following way:

A principal, i.e. a set of shareholders, engaging managers as agents to undertake entrepreneurial functions, wants to minimize the difference between the realized money equivalent of the principal's welfare and that money equivalent of the principal's maximum welfare, which could be achieved by actions of the agent.

For minimizing the difference, monitoring expenditures will be spent by the principal and bonding expenditures are imposed upon the agent.

How can the differences between the maximum welfare of the principal and the welfare realized by the agent be minimized?

For answering this question, the idea of agency costs is of no use because the residual loss cannot be found out (23, p. 553 ff.).

A manager as an agent should play the role of an entrepreneur in some markets on behalf of the principal. In this role the agent has to manage especially that kind of uncertainty which cannot be diversified or insured by contingent claims tradeable in Arrow-Debreu-markets (2; 9, p. 41 et al). That kind of uncertainty for which the abilities or functions of an entrepreneur are needed (i.e. "Misesian action", 15, p. 5-9) results from the fact that in general men (except for some entrepreneurs) are more clever at the end of any period than at the beginning. This implies that the plans of a lot of participants in some markets either do not include all states of the world which are necessary to build up rational expectations or the plans do not contain the credibility of some of the future states of the world or the amount of wealth in a reliable way. Credibility implies pre-ordinal as well as ordinal or quantitative personal probability (confer to these forms of uncertainty 11; 22 chapter A.III.).

To carry on the role of an entrepreneur in some markets is more uncertain than to play roulette; for the losses playing roulette can be "insured" in an Arrow-Debreu-world because the chances can be quantified. But entrepreneurial actions in markets try to realize actions and states of the world unforeseen in the decision trees of their competitors, customers or suppliers.

Indeed the idea of agency costs already breaks down in the simple case of a risky world, when a principal entrusts the task to play roulette on his behalf to an agent, i.e. if the table of roulette is regarded as a "market" for investments.

The agent who will maximize the welfare of the principal has to find out the right number in the next game. The gain would be 36 times the stake. Now, if the agent puts on a wrong number, is there a residual loss of 36 times the stake? Or does the residual loss amount to 32 times the stake, if the agent is so lucky to gain 4 times the stake? Is the residual loss determined by the expected value or some risk-averse expected utility-function of the principal?

This example leads us to the very question: Under what conditions would the welfare of the principal be maximized? Two solutions can be offered:

a) The decision the principal chooses by himself is regarded as the decision maximizing his welfare. But this proposal does not hold water: Managers as agents are ordered because they can solve special problems better than the principal. Therefore, the residual loss cannot be determined by the decision the principal would choose by himself.

b) The decision which an agent chooses solely acting in the principal's interest is regarded as the decision maximizing the welfare of the principal. But this proposal does not hold water either: How shall be proved under uncertainty and before the consequences are known which decision of the agent lies only in the interest of the principal?

Agents are of different quality, especially if their task is to play the role of an entrepreneur. Therefore, the question arises: What kind of agent would maximize entrepreneurial profit by what kind of action?

Such a question cannot be answered, because in acting as an entrepreneur the set of profitable actions is not known in advance. An entrepreneur must give solutions to unforeseen situations and make profits by innovations. For this task no probability distribution of gains and losses can be planned by a principal, before information has been gathered about the decisions of the agent in his role as entrepreneur. But the definition of a principal-agent-relationship is that the principal cannot observe the decisions of the agent.

Only under a very restricted set of conditions the idea of agency costs does not end in insolvable questions: when utility maximizing behaviour can be regarded as an empirical fact.

But in reality you can never be sure that an observable action maximizes utility. The aim of maximizing utility is not an observable fact, but only a methodological device or better: a metaphysical prejudice. The assumption of maximizing utility serves as a prerequisite for explaining relations between some observable facts as a result of situational logic. John Stuart Mill (17, 167-173) and Karl Popper (18, p. 97) have insisted on this.

Only if the plans of suppliers and demanders are in an ex ante competitive equilibrium and if this equilibrium could be observed in ex post reality, the statement should be justified that each person really maximizes its utility. Only a general equilibrium (if it is regarded as an observable fact) assures that allocative efficiency is fulfilled and utility is maximized for each principal and each agent.

But in general competitive equilibrium agency costs due to residual loss must be zero. This can be proved easily:

Competitive general equilibrium implies symmetric information. But the definition of the principal-agent-problem properly requires that the agent knows more than the principal. Therefore, the concept of a residual loss is incompatible with the neoclassic paradigm of utility maximizing in competitive equilibrium. Therefore we have to reduce agency costs to the sum of monitoring and bonding costs. But this leads to another objection: In general equilibrium theory all but one institutional arrangements are neglected. The only exception is the set of perfect and complete markets. Therefore, equilibrium theory cannot handle monitoring and other kinds of problems of an organizational framework.

In equilibrium nobody can gain abnormal returns by gathering information. Accounting, auditing and other kinds of monitoring are dispensable, if prices are the only signals of scarcity and by this, the best prediction of future prices (13). The implication of market valuation in equilibrium: "prices-are-signals", discloses a contradiction in the analysis of the monitoring and bonding expenditures as sources of agency costs by Jensen/Meckling (14, p. 316, 324). They assume that one dollar of current value of non-pecuniary benefits withdrawn from the firm by the managers reduces the market value of the firm by one dollar as well as one dollar of monitoring and bonding expenditures.

But if the market values an additional expenditure of one dollar as a reduction in market value of one dollar, then marginal cost equals marginal return by definition. Equilibrium is implied by this assumption and therefore the information of principals must be equivalent to that of the agents. The principal-agent problem does not exist anymore.

The real problem to solve would be: How do shareholders or "the market" value one additional dollar of monitoring or bonding expenditures? The agency costs approach can give no answer to this question.

3. The Flop Named "Transaction Costs"

There is a more general inconsistency in the several cost approaches to explain institutional arrangements as the separation of ownership and control or other principal-agent relations:

Either we look at the relationship between principals (stakeholders) and agents (managers) as a market relationship, then we have to determine the bundle of objects and services that constitutes the exchange ratio including the compensation incentives.

Or we regard the separation of ownership and control and other principal-agent-problems as a special kind of an authoritarian or hierarchical relation. In this case we need an economic theory of authority and organization which we only have at our disposal in nonquantitative terms (23, 546-551).

Transaction costs are used to measure the advantages and disadvantages between "markets" (or more market transactions = contracts for specified services) and "hierarchy" (or more authoritarian relations = unspecified or implicit contracts with some monitoring and incentive arrangements). The equilibrium theory and its offspring named transaction-costs-analysis neither give an explanation of any principal-agent-problem like the separation of ownership and control nor of the case of vertical integration, the original problem of the market-versus-hierarchy-debate (7, 27). This statement is based upon four reasons:

1. In equilibrium theory an alternative "market or hierarchy" does not exist. The coordination of economic activity takes place in different markets only. This first argument is well known and mentioned by Alchian/Demsetz (1, p. 777) and Cheung (6, p. 1o).

In the separation-of-ownership-and-control-debate we therefore first have to distinguish between the set of product markets and factor markets where managers have to act instead of the owners. The identity of ownership and control or the hierarchy between a sole owner and his manager has been substituted by a capital market separation between ownership and control.

In a market context the separation of ownership and control simply means that a new property right has been created to make one kind of specialization tradeable: that between agents as entrepreneurs and principals as capitalists.

2. The second reason is a consequence of market imperfections in reality. Transaction costs exist because contracts for each specified service of an agent cannot be written and enforced costlessly. Some organization costs occur in the case of unspecified or implicit contracts between the principal and an agent. But the comparison of transaction costs and organization costs will not determine the optimum amount of monitoring and bonding expenditure, if the price is a function of output, and therefore a residual claim results from the action of an agent. In imperfect markets it is simply not sufficient to compare costs, but to compare both alternatives by the sum of opportunity costs and residual claims.

   In imperfect markets there also exists another problem: transaction costs and production costs cannot clearly be distinguished. Only perfect competition guarantees the separation between transaction costs and production costs. In imperfect competition all costs finally are selling costs (16, p. 141-169). Transaction costs are selling costs in imperfect markets as well as all production costs. And in equilibrium theory the choice between those two markets remains either a black box or a purely technical problem determined by production function between the two markets.

3. The third reason is in close connection with my arguments against the agency cost approach.

   Transaction costs analysis is used as a yardstick to compare the alternatives "market" (or more specified contracts) and "hierarchy" (or more unspecified, implicit contracts).

But each yardstick must be independent of the alternatives to be measured. Yet costs are never independent of the "market". To determine costs implies a market as an institution because the price changes with the institutional facts and regulations in the special market. Transaction costs can be measured in three ways:

(a) Costs are defined as the sum of the mathematical products "price of each factor times input of this factor". But this definition implies that each "factor" including the services of an agent is valued in a market, and in this case markets and hierarchies cannot be alternatives.

(b) Costs are defined as opportunity costs. But if opportunity costs are understood as a quantitative term, they are only a mathematical implication of the maximizing procedure: "To cover costs and to maximize profits are essentially two ways of expressing the same phenomenon" (8, p. 1o8). Opportunity costs will become an observable fact and lead to testable hypotheses, only if a competitive equilibrium exists in reality (5, p. 85). But in this case market and hierarchy are no alternatives, and the principal-agent-problem is again defined away. This argument, used against the agency cost approach, leads to the conclusion: Agency costs can only be calculated, if a competitive equilibrium exists, because there is no need for monitoring; but in reality, where no competitive equilibrium exists and therefore there is a need for monitoring, the residual loss and the optimum amount of monitoring and bonding expenditures cannot be calculated. This follows simply from the fact that information values cannot be calculated if you do not know what the content of each information is.

(c) Adherents of the transaction cost approach may counter that "transaction costs" in their sense is not a quantitative term but only another name for an ordinal measurable "disutility". But an ordinal measurement and comparison of market (more disutility by specified contracts) and hierarchy (more disutility by implicit contracts for agency) implies the existence of an ordinal utility function over the two alternatives. The existence of such a "social welfare function" or "social decision function" integrating the interests of principal and agent cannot generally be warranted because of the well known "impossibility" problems (3, p. 25).

4. The fourth reason results from the very task that a manager as an agent should play the role of an entrepreneur on behalf of the principal and the principal controls by non-market alternatives.

4. Alternatives to Incentives as Solution of Principal-Agent-Problems

To treat principal-agent problems with the tool of agency costs narrows the view of scientists for the solutions found in practice and methodologically leads them into desert. Fee schedules are only one way to approach a solution, penalty (sanctions) is another one, and accounting free of manipulation is a third way for the principals to reduce their lack of information and to improve the use of their contractual rights or the exit from contracts with an agent.

The first alternative to incentives in the relationship between principal and agent is strict orders with all their consequences, if these orders are not obeyed against better reason. The drama "Prince of Homburg" by Heinrich v. Kleist is based on this conflict.

Going over from military to economic action on somebodies behalf and instructions, especially to earning money by investment for others, even here one way of solution is strict liability at the risk of his neck for the agent who invests the money of the principal. But in contrast to strict orders in a military principal-agent-relationship the agent is left a far reaching liberty to adapt his actions to changing circumstances, but certainly only if the money invested by the principal can be separated from the rest of his property. The legal institution of limited liability is one of the prerequisites which are necessary for the development of today's principal-agent-relationships in capital markets, and anyway even for the existence of efficient capital markets, because a tendency to complete capital markets implies insurance against the risk of losses and insurance requires a limitation of the maximum loss.

In Roman Law the partners in a company had unlimited liability. But this did not hamper the desire for acquisition of limited liability and the rudimentary development of capital markets (23, p. 351).

If a Roman capitalist wanted to venture a part of his money into an enterprise he went to a slave-trader and picked out a person he held competent to manage his business. This slave had to hold in trust the invested money as a special good (peculium). If the enterprise went wrong only the limited capital of the peculium was lost. And if in this case the capitalist was not satisfied with the management of the slave he could leave him in the hands of the creditors or he could do what another Roman did with his personal philosopher: fatten the fish with his flesh.

Today these solutions in principal-agent-relationships are not possible any more. But perhaps the principal-agent-problems in capital market relationships of today, i.e. the total scope of protection of creditors, the opening of bankruptcy proceedings etc., are so difficult, because our society avoids radical sanctions to support the fulfilment of duties from contractual relations. Debtors' prison is abolished and the punishment for bankruptcy with losses of billions of pounds is modest compared for instance with the ordonnance de commerce of Louis XIV of France: death penalty. Even the losses of property can nowadays be confined by the bankrupt: A banker with unlimited liability may for instance sell his property to his wife shortly before he has to report the bankruptcy of his firm while in return his wife gives him the now valueless shares in the bank, which she inherited from her father and flees from the country.

The social feasibility of sanctions against the agent and his failure as an entrepreneur or his moral hazard have extremely been restricted by modern law. This is because principal-agent-relations are fundamental questions of the distribution of power in labour-relationships. And since Roman Law the development in labour legislation has been more and more to blame the owner of capital for the failure of his agent. This exploitation of capital by (I stress) qualified work, i.e. the trend of the exploitation of capital by functionaries and managers, is only burred by a legal safeguard in form of entitlements which connect capital market facts and labour market facts, especially the voting rights to elect the board of directors.

By the entitlements it is attempted to approximate the fetish of the unity of ownership and control. The unity of ownership and control rules out every principal-agent-problem by definition.

The fetish of the unity of individual ownership and control was built up into the indispensible mark of a liberal and capitalist economy by the so called "classic" theory of the firm. This ended up into a misjudgement of what is needed to allocate the resources of a market economy by prices:

If we strive for economic reasons to justify certain legal maxims and rules in market or labour relations under the general objective of allocative efficiency and if we explicate this objective by the general (Walrasian) equilibrium the unity of ownership and control is ruled out by this scientific approach. For given allocative efficiency under the conditions of uncertainty in an Arrow/Debreu-type model the economic agents as consumers own all shares of the existing firms as productive economic units. But the proof of the existence of a general Walrasian equilibrium requires an empty set of intersections between consumers and producers. Here personal ownership and the control are strictly separated by perfect and complete capital markets with regard to the diversification and insurance of risks and by perfect markets for the services of entrepreneurs or managers.

From this it can be derived: The so called "classic" firm with unity of ownership and control, as it supposedly existed in the 19. Century (4, p. 8 f.; 25, p. 134-136), is <u>not</u> an integral part of an allocative efficient competitive economy. On the contrary: The unity of ownership and control <u>denies</u> the efficiency of capital markets.

As the unity of ownership and control is by definition contradictory to allocative efficiency joining together voting rights for contracts in labour markets (e.g. the right to elect the board of directors in companies) and voting rights for the withdrawal of capital (e.g. dividends but also the options on an new share issue) stands against the improvement of allocative efficiency by the market process. Labour markets have just to be regulated and deregulated on other grounds as financial markets. This concerns especially those labour markets where the provided services mainly consist of practising entrepreneurial functions, as it is the case in markets for managers.

5. Conclusion

A theory of agency costs helps to solve principal-agent-problems if and only if all questions of labour legislation can be put down to considerations of costs. But if they could there would exist no

problem of hierarchy. Hierarchy follows by economic reasons just from a superiority of knowledge, e.g. in a principal-agent-relationship. To handle principal-agent-problems by considerations of costs incorporates an inherent contradiction if costs are understood as quantitative terms.
From this contradictions some authors try to save themselves by making the following excuses: Agency costs are neither payments nor opportunity costs in the theoretical sense of the London School of Economics (8, p. 5) but only a metaphorical expression. But this does neither save the agency-cost-approach nor the similar transaction-cost-approach from failure as a basis for theory-building. Metaphors as "agency costs" used in a pre-scientific sense outside a scientific meaning of the term can at most lead to a theory-drivel but not to a theory. Therefore my concluding thesis is: A basic theoretical concept (here: agency costs) used as a metaphor does only verify the incompetence of some scientists to speak clearly of their supposed problem and to get hold of it with their tools at hand. Cost is a quantitative concept. That is not only a question of understanding each other in a scientific community. It is also a question of the style in developing economics as a science. To improve the style, Nietzsche says, is to improve the thought - and nothing else!

References:

1. Alchian, Armen A.; Demsetz, Harold: Production, Information Costs, and Economic Organization. In: American Economic Review 62 (1972), 777-795.

2. Arrow, Kenneth J.: The Role of Securities in the Optimal Allocation of Risk-bearing. In: Review of Economic Studies 31 (1964), 91-96.

3. Arrow, Kenneth J.: Social Choice and Individual Values. 2nd ed. New York-London-Sydney 1963.

4. Berle jr., Adolf A., and Means, Gardiner C.: The Modern Corporation and Private Property. New York 1932.

5. Buchanan, James M.: Cost and Choice. Chicago 1969.

6. Cheung, Steven N.S.: The Contractual Nature of the Firm. In: Journal of Law and Economics 26 (1983), 1-21.

7. Coase, R.H.: The Nature of the Firm. In: Economica, New Series, IV (1937), 386-4o5.

8. Coase, R.H.: Business Organization and the Accountant (1983), reprinted in: L.S.E. Essays an Cost, edited by J.M. Buchanan and G.F. Thirlby. London 1973: London School of Economics and Political Science, 95-132.

9. Debreu, Gérard: Theory of Value. New York-London-Sydney 1959.

1o. Fama, Eugene F.; Jensen, Michael C.: Agency Problems and Residual Claims. In: Journal of Law and Economics 26 (1983), 327-349.

11. Fine, Terrence L.: Theories of Probability. New York-San Francisco-London 1973.

12. Grossman, Sanford J.; Hart, Oliver D.: An Analysis of the Principal-Agent Problem. In: Econometrica 51 (1983), 7-45.

13. Hayek, F.A. von: The Use of Knowledge in Society. In: American Economic Review 35 (1945), 519-53o.

14. Jensen, Michael C.; Meckling, William H.: Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. In: Journal of Financial Economics 3 (1976), 3o5-36o.

15. Kirzner, Israel M.: Perception, Opportunity, and Profit. Chicago-London 1979.

16. Kirzner, Israel M.: Competition and Entrepreneurship. Chicago-London 1973: University of Chicago Press.

17. Mill, John Stuart: Essays on Some Unsettled Questions of Political Economy. London 1844. German translation Frankfurt 1976 cited.

18. Popper, K. R.: The Open Society and its Enemies. 5th ed. London 1966.

19. Radner, Roy: Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship. In: Econometrica 49 (1981), 1127-1148.

2o. Rees, Ray: The theory of Principal and Agent. Part I. In: Bulletin of Economic Research 37 (1985), 3-26, Part 2, 75-95.

21. Roscher, Wilhelm: Die Grundlagen der Nationalökonomie. Stuttgart-Tübingen 1854.

22. Schneider, Dieter: Investition und Finanzierung. 5th ed. Wiesbaden 198o.

23. Schneider, Dieter: Allgemeine Betriebswirtschaftslehre. 2nd ed. of Geschichte betriebswirtschaftlicher Theorie. München-Wien 1985.

24. Schumpeter,J.A.: Unternehmer.In: Handwörterbuch der Staatswissenschaften. 4th ed., Vol. 8, Jena 1928, 476-487.

25. Sen, Armartiya K.: Collective Choice and Social Welfare. San Francisco etc. 197o.

26. Stiglitz, Joseph E.: Credit Markets and Control of Capital. In: Journal of Money, Credit, and Banking 17 (1985), 133-152.

27. Williamson, Oliver E.: Markets and Hierarchies: Analysis and Antitrust Implications. New York 1975.

Agency Costs are not a "Flop"!*)

Reinhard H. Schmidt

Summary: The following comment on the paper by Dieter Schneider in this volume only deals with his critique of the agency cost approach. He considers the agency cost approach to be "a flop". It is demonstrated that his fundamental objection, i.e. that agency costs and, in particular, the "residual loss" are not measurable, does not justify this scathing verdict. In general the role of measurability of costs depends on how the cost concept is used. Therefore, Schneider's views of the explanatory function and the pragmatic function of the agency cost concept are analyzed. As his notion of explanation seems to be too restrictive and as his understanding of how the agency cost concept should help to solve the principal's problem is inadequate, his harsh critique of the entire approach has to be rejected. The concept of agency costs may be "metaphorical", but that does not reduce its value as a theoretical tool.

Table of Contents

---

*) Valuable criticism and suggestions from Gabriel Schor and Eva Terberger are gratefully acknowledged.

Agency Theory, Information, and Incentives
ed. by G. Bamberg and K. Spremann

1. Introduction

As are most of Dieter Schneider's recent papers assessing and attacking neoclassical economic and financial theory[1], the paper in this volume is fascinating, stimulating and provocative. At the same time, however, it is not easy to read because of the plethora of sophisticated arguments which are intertwined in a subtle way and which are more hinted at than developed in depth in this short essay. There are three points in Dieter Schneider's critique of the agency cost approach which I find most interesting and which may be regarded as his central objections to this approach:

(1) Dieter Schneider severely attacks the use of the concept of costs in the agency cost and transaction cost approaches as being too lax and, indeed, logically inconsistent.

(2) He regards the agency cost approach as having failed to fulfill its explanatory function, as he sees it. This judgement may be due to his view that

(3) the agency cost concept is inadequate to fulfill a practical or normative function which he believes the approach claims to have.

These three fundamental objections are a reflection of a basic underlying position which is, in my opinion, grounded in a dogmatic view of cost as a quantitative concept, and implicit - but not necessarily adequate - notions of what characterizes good explanations and good "solutions", and a very specific interpretation of "the agency problem".

I will only discuss his main objections (1) to (3) to the agency cost approach. Because of the limited space available here and because I also have my objections to the transaction cost approach, I do not want to discuss the paper's critique of the latter. I shall also only touch briefly his "Austrian style" alternative to the agency cost approach.

---

[1] The references 14 (pp 459-569), 15, 16 and 17 are only a selection.

2. Cost as a Quantitative Term

For Dieter Schneider, cost has to be a quantitative and observable term. In his pioneering "Geschichte betriebswirtschaftlicher Theorie" (14, p. 389 ff) he very clearly demonstrated that cost in the sense of opportunity cost is not generally observable and that observable costs based on observable prices are not generally relevant for decision making. In order to be of practical relevance and, at the same time, observable, i.e. capable of being measured in an empirically meaningful way, any concept of cost has to be based on (observable) prices which are independent of any decision which might result from the outcome of the measurement of costs. Prices are independent of decisions in General Competitive Equilibrium (Dieter Schneider seems to believe that this sufficient condition is also necessary). Anyone who uses the term "cost" in a positive theory must, according to Dieter Schneider, make the implicit assumption that a General Competitive Equilibrium, as it is analyzed in the relevant theories of Debreu (1) and others, is a well-defined concept and a fact of life: If you use the term "cost", you either make a statement about reality or you merely produce meaningless verbiage. This is the line of reasoning which makes him call agency costs and transaction costs "offsprings" (p. 487) of General Equilibrium Theory and which makes him apply his objections to General Equilibrium Theory as a theoretical construct and as a statement about reality (15) to prices in general, to all concepts of cost and to the agency costs concept. When General Competitive Equilibrium is not given, prices - and thus costs based on observable prices - are not independent of decisions and are, therefore, not an appropriate "measuring instrument". As I understand his paper, this is why he calls the concept of agency costs "metaphorical" - a word which he uses in a pejorative sense.

To his general critique of all cost concepts, Dieter Schneider adds a special - and, in my view, stronger - argument against the inclusion of any concept of cost in an approach which, like the agency cost approach or the transaction cost approach, aims at explaining institutions, because such cost concepts can be regarded as inconsistent or even contradictory: He regards the assumption that General Equilibrium prevails, which he deems implicit in any use of the concept of cost, as equivalent to considering General Equilibrium Theory as a true statement about reality. But this theory implies, in the final analysis, that there are no institutions which could be analyzed, with the exception of property and a system of markets (p. 486). This creates the inconsistency between the object which is to be analyzed and the instrument of analysis.

His statement about the importance of General Equilibrium and separability for a concept of cost as a quantitative term - as he defines it - is valid. And his plea for using concepts of measurement which are theoretically well founded, independent of the object to be measured, and empirically applicable, is unobjectionable in abstracto. But are this statement and this plea relevant for the evaluation of the agency cost approach? Why should issues of measurement be so important in this case? I fail to see why observability should be tantamount to measurability (14, p. 57). Is something for which we do not have a satisfactory theory of measurement simply unobservable or even non-existent? Is it generally impermissible to use concepts which can be well explicated (only) within a theory which may have some drawbacks? Dieter Schneider rightly points out difficulties of the cost concept per se, but this does not necessarily mean that it should be discarded.

Even if his general critique is accepted as valid and even if it is admitted that cost is, in a certain sense, "metaphorical", the relevance of the difficulties arising from the "metaphorical" concept of cost depends on how this concept is used. His critique is unnecessarily sweeping. All of the contributions to - and applications of - agency cost theory of which I am aware employ costs or prices without bothering about whether they are (General) equilibrium prices or even observable. It has yet to be shown that this laxity has any negative consequences. Dieter Schneider does not even attempt to demonstrate that this is so. It would have been more relevant if he had argued that the notion of agency costs - not agency theory in general - is not useful and if he had claimed that its limited usefulness is due to the problematic nature and to the metaphorical character of the basic concept which he pointed out. But this does not seem to be Dieter Schneider's argument. Instead of usefulness he focuses on logical consistency or it seems that he equates the one with the other.

The essential reason, however, why his insistence on measurement is questionable is that, according to the definition of agency costs, they are never observable. Whether this matters depends on, both, the notion of explanation which is used and on the understanding of "the agency problem". This will be elaborated below.

"Agency costs" may indeed be a metaphor. But this is not necessarily a weakness[2]. Indeed, one could consider agency costs to be a good metaphor because, as such, it directs attention to the problem of taking into account and assessing the consequences which an asymmetrical distribution of information can have for the way people organize their cooperation. The importance of quantitative terms for metaphors - or more generally: the role of measurability for good theories and even the function of theories - is not as clear as Dieter Schneider suggests. His main critique of the approach, i.e. that agency costs are not measurable, is thus unconvincing for two reasons: for one thing, agency costs are by definition not even observable, and for another, it is imperative to look how the agency cost concept is used in scientific practice. As this task would be beyond the scope of this brief comment I shall only deal with the uses of the agency cost concept that Dieter Schneider discusses.

3. Do Agency Costs Fail to Explain Something?

Dieter Schneider states in the first sentence of his summary that "agency costs ... are generally used to explain agency problems" (p.481).Throughout the paper he makes it clear that, in his view, this explanatory function is not fulfilled. But nowhere in his paper does he clarify

---

[2] A positive assessment of theories as metaphors is discussed in 4, pp 111 ff; see also 13.

(1) what object or phenomenon agency costs are supposed to explain,
(2) in what sense they might explain something, and
(3) what requirements an explanation has to meet in order to be regarded as "good" or "satisfactory".

Question (3) is the easiest one to answer: In Dieter Schneider's view, a good explanation should be a deductive nomological explanation centered around an unfalsified, law-like statement. However, in the theoretical and applied agency literature such explanations are rarely found and not even sought. Instead, merely theoretical explanations (e.g. in 5 and 17) or functional explanations (e.g. in 3 and 7) are provided. The latter have the following structure : Organizational form X can be observed because it survives; its survival is due to its ability to cope better than other, competing forms of organization with the difficulties caused by asymmetrically distributed information. In short: it survives because it exhibits lower agency costs. It would, however, be wrong to read this statement as a substantive explanation which says that agency costs "cause" institutions. It is (only) a meta-theoretical statement: Look at the things which are hinted at by the term "agency costs" if you want to develop a good explanation.

The law-like statement in functional explanations is the assumption of rational behaviour of all individuals pursuing their aims under given circumstances. But this assumption is not considered to be true and/or testable. The distinguishing mark of a good functional explanation is, rather, that the phenomenon to be explained follows logically from a straightforward description of the relevant situation and the assumption of rationality. Dieter Schneider would not reject this type of explanation (p.486) but his insistence on observability in the strict sense of measurability is not, in my view, compatible with this notion of explanation.

But then the question arises: In what sense could something called "agency costs" be meant to explain an object called "agency problems"? The literature on agencies tries to characterize a type of situation, the agency situation. It results from the interaction of people having different information, different sets of acts to choose from and some common and some divergent aims. The welfare of one part, generally called the principal, is in some way influenced by the decisions of the other part, called the agent. Often, but not necessarily always, the principal has an opportunity to influence the agent's choice of an act. But this control is not complete and/or not without its costs. In some cases the principal can give orders and supervise (imperfectly) the agent's behaviour, or he may be in a position to institute a sharing rule for pay-offs, i.e. an incentive scheme.

An agency situation is theoretically trivial as long as there is not, to some extent, a divergence of interests resulting in an incentive problem, and as long as information is not distributed asymmetrically and the transfer of information and the writing and enforcement of contracts is not difficult or costly. Where information distribution is symmetrical and/or information is freely available and contracts can be written and enforced costlessly, the incentive problem could be eliminated by writing the necessary contracts specifying what the agent has to do (or must not do).

When interaction or cooperation involves an incentive/information problem, this may be called an "agency problem" for the simple reason that in this situation the maximum welfare achievable is less, in a Pareto-sense, than in the ideal world of perfectly and costlessly enforceable contracts and/or symmetrical information ("first best optimum"). Only as a limiting case the first best optimum may also be achievable. The achievable optimum is, therefore, weakly Pareto-inferior to the unconstrained, or cooperative optimum. If it is strictly inferior it is a "second best optimum".

Some activities that are designed to achieve an "improvement in second best", like bonding of the agent and monitoring by the principal, may be feasible and rational in such a situation. But these activities will be imperfect and costly, so that the "first best optimum" can (normally) not be attained.

One may call the final difference in welfare, expressed in monetary units, the "agency costs" and divide them up into the welfare loss resulting from a different action choice - compared with the "first best" case - (the "residual loss") and the expenditures for devices to reduce the welfare loss (the "bonding and monitoring costs"). It follows from their definition that the agency costs and, in particular, the residual loss are not observable. The "first best optimum" is (in general) not a real, observable state of affairs! This, of course, does not imply that in a theoretical model agency costs could not be clearly defined and that in an empirical study agency costs could not be approximated. Or is it not an example of a rough approximation if in a specific cooperative savings scheme in Africa about ten percent of the funds are reported to be misappropriated every year (on this example see 12)?

The agency situation is a general concept or a formal structure. Real situations like the relationship between shareholders and managers or masters and their servants may be interpreted as principal-agent situations. But they are just examples or applications. There is not <u>the</u> right or "wrong idea about the facts to explain"(p.481).

I can now directly address the alleged explanatory function of agency costs. It does not seem that agency costs "explain" - in any meaningful sense - the problems inherent in the general structure of the agency situation or in any real situation which may be interpreted as an agency situation. When cooperation becomes difficult, this is "caused" by the incentive/information problem - assuming it can be said to have a cause at all. "Agency costs" are only symptoms. They do not explain the welfare loss nor the attempts to reduce it by "bonding and monitoring".

The disadvantages (or costs) of any specific device to mitigate the adverse effects of the incentive/information problem determine to what extent this device should be employed. Accordingly, one could say that the <u>extent</u> of "bonding and monitoring" in a <u>specific</u> case might be explained by "bonding and monitoring costs". But this does not seem to be what Dieter Schneider means by his statement. Any more general claim that "agency costs" <u>should</u> explain - and fail to explain - agency problems would appear to be unfounded.

Nevertheless, agency costs are by no means useless in explaining real phenomena. Explaining real phenomena in the light of agency theory requires that one apply knowledge about a formal structure ("the agency situation") to a specific case. This interpretation of reality may be facilitated by a knowledge of the intuitive (metaphorical) concept of agency costs and of the formal properties of the structure called the "agency situation": The agency cost concept directs our attention to the incentive/information problem and its consequences and to the difficulties associated with all devices that are employed to cope with these consequences. For this - in my view very important - heuristic function the subtle problems in empirically measuring agency costs - and, in particular, the residual loss - are largely irrelevant. Discarding the concept of residual loss, which Dieter Schneider attacks most vehemently, would deprive the agency concept of its heuristic function and of much of its power. Apart from this - and apart from the function of communicating the problems and results of an investigation - the concept of agency costs may, however, be considered less important. Any detailed analysis of institutional arrangements will have to deal directly with the incentive/information problem. It cannot restrict itself to comparing institutions by means of a global - and non-operational - measure of "the agency costs". And in theoretical analyses the concept of agency costs is, quite rightly, not used at all.[3)]

4. Do Agency Costs Fail to Solve a Practical Problem?

Dieter Schneider also investigates the contribution of the agency cost concept to the solution of practical problems in an agency situation. And as one might expect, he finds the concept useless because it is not possible to quantify the agency costs - and in particular the residual loss component - in practical decisions.

Dieter Schneider discusses a special agency situation or problem, namely the selection, hiring and "using" of an agent as a manager in the role of an entrepreneur. However, I fail to see why the delegation problem in his first example of the agent playing roulette on behalf of the principal should be an agency situation. There is no incentive

---

3) This applies equally to papers like 5, 10 and 17 which analyze the formal structure of the principal-agent relationship and to papers like 8 which use this structure to investigate issues in financial theory.

problem and no asymmetrically distributed information in this "simple case of a risky world" (p. 484). Moreover, the frequent references to "Misesian action", "true uncertainty" and the "entrepreneurial role" tend, in my view, to obscure the crucial point: Does the inability to transfer directly a specific model of agency theory with its formal apparatus to the real world imply that the basic idea is necessarily inapplicable? This cannot be taken for granted as long as the quantification of the residual loss is not regarded as the central point. But for Dieter Schneider it is the central point.

Most of the probing questions on p. 485 of his paper are of an applied normative character. They refer to "the principal's problem" (10). It is not quite clear to me whether these normative issues are discussed in the framework of positive theory, namely in terms of analyzing how people in fact behave (rationally) in a given situation, or whether there is a genuine interest in the practical application of the agency cost concept as a management tool. I am not aware of any attempt in the literature to advocate this kind of practical application.

Dieter Schneider's argument seems to be the following: In his view, the agency cost approach solves the practical problem if and only if a principal (an owner) could select the optimal agent (manager) and the optimal contractual arrangement for the principal-agent relationship by employing the exact quantity of (total) agency costs as a yardstick for gauging alternatives. And only if this were the case would he be prepared to accept the agency cost approach in positive, explanatory theories: Institutions are a consequence of agency costs in the sense that people select institutional arrangements on the basis of the criterion of (total) agency costs.

I find two reasons in the paper which attempt to explain why the agency cost approach does not solve the principal's problem. Both amount to saying that it is infeasible, in principle and in practice, to calculate (total) agency costs and, thus, to use them as a criterion. One is related to the issue of "true uncertainty", "entrepreneurship" etc. briefly touched on above.

His second argument is, in my view, more important. It has to do with the alleged circular reasoning in the decision-theoretic treatment of the value of delegation (e.g. 2, 6, 9). Dieter Schneider is here referring to the old argument to the effect that one cannot assess the value of a given piece of information without first knowing what it is or that one cannot find the optimal simplification of a decision model for one single decision. But the intellectual appeal of this seeming paradox is lost as soon as the value of not just one piece of information, but rather that of the use of an information system, is to be assessed and as soon as a simplification for a series of decisions is sought (11).

The decision-theoretic evaluation of delegation - which Dieter Schneider equates too directly with the agency problem - may, indeed, be beset with the problem of logical circularity. But this is the case <u>if and only if</u> the agent is not considered by the principal to possess better information or other advantages and if maximization (of the value of delegation) in a very literal sense is attempted. In this very specific case, the best decision the agent could make would be the one which the principal could make himself by utilizing the information he needs to have in order to evaluate the agent. Therefore, there is no advantage in hiring an agent and delegating decision making competence to him.

The problem of delegation is, however, ill defined whenever the principal could decide as well as the agent; then, by implication, the evaluation of delegation is logically circular. But if, for instance, the agent is better informed, employing him would amount to using an information <u>system</u>[4]. And in this case - as well as in other cases in which the agent has a comparative advantage - the logical circle is avoided and the principal's own optimal decision can be worse, in terms of expected utility, than the decision which the agent is expected to make. An economically meaningful interpretation of the delegation problem implies a trade-off, in the principal's view, of some advantage of delegation against the disadvantage that the agent will not necessarily decide in exactly the same way as the principal would

4) In 9, Laux treats the delegation problem as an application of the formal structure of assessing an information system. Reasons why a well-informed principal may want to use an agent could be that the agent is incapable of communicating his orders (e.g. Al Capone in jail may have to create an incentive system for his fellow bandits, because he is not allowed to tell them what he knows best).

if he had the agent's knowledge or other comparative strength. But it has to be emphasized that this comparison involved in assessing the disadvantage of delegation is a hypothetical one: Except for the strange case of circularity, the principal does not have (e.g.) the knowledge of the agent and could not make the "really" optimal decision. The practically relevant comparison is, instead, that between the prinicpal's own - e.g. poorly informed - decision and the expected decision of the - e.g. better informed - agent.

The formal theory of agencies makes this point sufficiently clear: It adds a constraint which assures incentive compatibility to the cooperative maximization problem. The "guesses"[5] as to what decisions the agent could take are replaced by the - calculated but ex-ante uncertain - solutions of the agent's maximization problem.

In the formal agency theory one learns to distinguish between, on the one hand, comparisons of first best and second best, and on the other hand, improvements in second best. The principal's choice is between different attainable situations, e.g. with more or less bonding and monitoring activities. The first best or cooperative solution may be among the attainable situations. In this limiting case it will be chosen. In general, however, the first best optimum is unattainable and therefore irrelevant for the principal, and it is, in most cases, also unknown to him. Not a principal, but only the (theoretical) researcher can, in a model situation or with the advantage of hindsight, undertake an exact comparison of first best and second best optima. It is simply beside the point to say, as Dieter Schneider does, that the principal cannot solve his problem by calculating (exactly) the residual loss: Of course he cannot, otherwise he would not have his problem. The relevant decision problem of the principal can be solved, and the exact or theoretical solution to this problem is independent of the first best optimum and, in particular, of its exact quantification.

In practice or real life, the decision to hire an agent and to design contracts and/or incentive schemes is, of course, difficult. Limited

---

[5] The probability assessments of the principal concerning the state-dependent action choice of the agent in Laux' treatment of the delegation problem (9, pp 69-77) are exogenous to the model.

capacity to acquire and process information - or the "true uncertainty" of an Austrian type - forces people to settle for approximations instead of strict maximizations. Nevertheless, what they approximate is not the irrelevant or hypothetical comparison with the fictitious first best optimum but the difference between alternative attainable situations. This implies that even if the agent is engaged because he knows more, and even if he can pursue his own interests to some extent, a principal will not be completely incapable of making a "wise" decision as to which agent to select and how to arrange the contractual relation. Such decisions are made many times every day when people are hired, doctors or lawyers are selected and academics are invited to write papers.

However, one reservation is in order here: The first best optimum is practically irrelevant only if different second best alternatives are evaluated, provided that a list of such options has been compiled. In order to find activities of a bonding and monitoring type, which are then to be evaluated, it may be a useful heuristic to think of a - necessarily vaguely described - first best optimum: In preparing the decision which is intended to solve his practical problem, a principal will, most probably, try to understand his problem and, therefore, consider the consequences of divergent interests in a setting of asymmetrical information in comparison to an ideal or fictitious situation. This implies that he will try to estimate, at least implicitly, the total ageny costs.

All these considerations lead to the conclusion that a principal[6] will act as if he were minimizing total agency costs. His inability to do so in a very special situation - where the verdict of logical circularity applies - and the general theoretical irrelevance of the "first best optimum" do not render this as-if statement wrong or vacuous or paradoxical.

---

6) The problem can be cast in the form of the maximization of either the principal's or the agent's utility - with the other one receiving a "reservation utility" (see, e.g. 5).

5. Implications

The implications of my characterization of the normative or decision making aspect of the principal-agent problem for the explanatory function are straightforward. As the solution of the practical agency problem does not require a comparison with a theoretically defined first best and, thus, does not require an exact quantification of the residual loss or of total agency costs, one cannot expect to observe principals calculating these costs. Quite apart from the problems of observing maximizing behaviour - which Dieter Schneider would readily concede - the measurability of agency costs (or of the first best optimum) is not a crucial issue. Therefore the examples ("Musterbeispiele") of a successful application of the formal structure of agency theory cannot, by definition, be empirical examples of people minimizing total agency costs. The inability to measure agency costs exactly does not reduce the value of the formal theoretical idea ("Lösungsidee") [7]. Neither does the limited - in fact, only heuristic - role of agency costs in decision making reduce the usefulness of the agency cost concept in positive theories.

I tend to believe that agency costs are not the most valuable part of agency theory. But it seems unfounded and unfair to call it a flop - and, to add just one final remark, it is equally unfair to place the concept of agency costs on the same level as the extremely fuzzy concept of transaction costs.

References

1. Debreu, G. (1959): The Theory of Value, New York - London

2. Demski, J.S. (1976): Uncertainty and Evaluation Based on Controllable Performance, in: Journal of Accounting Research, Vol. 14, pp 230-245

3. Fama, E.F., Jensen M.C. (1983): Separation of Ownership and Control, in: Journal of Law and Economics, Vol. 26, pp 327-349

---

7) See 14, pp 53 - 62, on the relationship between theoretical ideas and (empirical) examples in good theories.

4. Hesse, M. (1980): Revolutions and Reconstructions in the Philosophy of Science, Brighton

5. Holmström, B. (1979): Moral Hazard and Observability, in: The Bell Journal of Economics, Vol. 19, pp 74-91

6. Horst, M., Schmidt, R.H., Terberger, E. (1982): Risikoaufteilung durch Prämiensysteme, in: Zeitschrift für Betriebswirtschaft, Vol. 52, pp 942-958

7. Jensen, M.C., Meckling, W.H. (1976): Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, in: Journal of Financial Economics, Vol. 3, pp 305-360

8. Krahnen, J., Schmidt, R.H., Terberger, E. (1985): Der ökonomische Wert von Flexibilität und Bindung, in: Ballwieser, W., Berger, K.H. (eds): Information und Wirtschaftlichkeit, Wiesbaden, pp 253-285

9. Laux, H. (1979): Grundfragen der Organisation, Berlin-Heidelberg-New York

10. Ross, S.A. (1973): The Economic Theory of Agency: The Principal's Problem, in: The American Economic Review, Vol. 63, pp 134-139

11. Schmidt, R.H.(1983): Zur Lösbarkeit von Vorentscheidungsproblemen, in: Bühler, W. et al. (eds): OR-Proceedings 1982, Berlin - Heidelberg - New York, pp 417-426

12. Schmidt, R.H. (1986): Credit Supply, Self-Help and the Survival of Financial Institutions in Developing Countries, in: Herder-Dorneich, Ph. (ed.): Jahrbuch für Neue Politische Ökonomie, 5. Band, Tübingen, pp 262-279

13. Schmidt, R.H., Schor, G. (1987): Modell und Erklärung in den Wirtschaftswissenschaften, in: dies. (eds): Modelle in der Betriebswirtschaftslehre, Wiesbaden, pp 9-36

14. Schneider, D. (1985): Allgemeine Betriebswirtschaftslehre, 2nd.ed.of Geschichte betriebswirtschaftlicher Theorie, München - Wien

15. Schneider, D. (1986): Aufstieg und Niedergang eines Forschungsprogramms: allgemeine Gleichgewichtsanalyse, manuscript, to be published in: Studien zur Entwicklung des ökonomischen Denkens VII (Berlin 1987)

16. Schneider, D. (1987): Modellplatonismus in der Kapitalmarktgleichgewichtstheorie, in: Schmidt, R.H., Schor, G. (eds): Modelle in der Betriebswirtschaftslehre, Wiesbaden, pp 159-196

17. Shavell, S. (1979): Risk Sharing and Incentives in the Principal and Agent Relationship, The Bell Journal of Economics, Vol. 10, pp 55-73

About Contributors

WOLFGANG BALLWIESER (* 1948) is professor of accounting and auditing at the Universität Hannover. He got three degrees from the Johann Wolfgang Goethe Universität in Frankfurt am Main: Diplom-Kaufmann 1974, Dr. rer. pol. 1977, Habilitation in business administration 1981. Since 1982 he is full professor in Hannover. He was visiting professor at the Universität Wien 1983/84. His publications deal with various topics of accounting, auditing, commercial and capital market law, the valuation of the firm and liquidity management.

GÜNTER BAMBERG (* 1940) studied mathematics at the Universities of Saarbrücken and Bonn. He got his diploma degree at the University of Saarbrücken in 1966, his doctoral degree at the same University in 1968 (thesis on game theory), and his Habilitation at the Technical University of Karlsruhe in 1970. Since 1970 he is full professor of statistics, University of Augsburg, where he stayed despite calls to Osnabrück and Heidelberg. He is co-editor of the Statistical Papers and Mathematical Systems in Economics. His books and articles deal with statistics, econometrics, Operations Research, and decision theory.

MARTIN J. BECKMANN studied mathematics at the University of Göttingen 1945 - 1948 and economics at the University of Freiburg i.B. 1948 - 1950, receiving his doctorate from Walter Eucken with a thesis on "Distribution under Alternative Economic Systems". He was a Post-doctoral Fellow at the University of Chicago 1950 - 1951 and a Research Associate of the Cowles Commission for Research in Economics, the University of Chicago 1951 - 1955 working under T.C. Koopmans and J. Marschak. He moved with the Cowles Foundation to Yale University (1956 - 1959). Since 1959 he has been teaching economics (mathematical economics, managerial economics, economics of organizations) at Brown University, Providence R.I. Concurrently he was Professor of economics at the University of Bonn (1962 - 1969) and is Professor of Applied Mathematics at the Technical University of Munich. His research interests are in the economics of transportation and location, dynamic programming, and the economics of organizations.

HELMUT BESTER (* 1953) is assistant at the Department of Economics, University of Bonn. He studied economics in Bonn (Diplom-Volkswirt 1979) and at the London School of Economics (M.Sc. 1981). He received his doctoral degree (1984) and his Habilitation (1987) from the University of Bonn. His research interests include the economics of uncertainty and information and non-cooperative bargaining theory.

MARINA BLICKLE (* 1960) studied mathematics and economics at the University of Ulm. She received her Diplom-Wirtschaftsmathematiker degree at the University of Ulm in 1985. Since 1986 she has a research and teaching position at the Department of Economics, University of Ulm. Her current interests of research are agency theory and information economics.

RALF EWERT (* 1957) is currently working as Akademischer Rat at the University of Passau. He first studied music at the Robert-Schumann-Institut, Düsseldorf, and then business administration at the University of Cologne, from which he received the "Diplom-Kaufmann"-degree in 1982. He got his "Dr. rer. pol."-degree from the University of Passau (1986), and his doctoral thesis deals with the application of the financial theory of agency in the field of external accounting. He has published several papers on decision theory (jointly with Thomas Schildbach) and finance. His current research interests are in the above mentioned areas as well as in the fields of managerial accounting, information economics and the economic analysis of law.

GUSTAV FEICHTINGER (*1940) is full professor for operations research at the Technical University of Vienna. He studied mathematics and physics at the University of Vienna, where he got his doctoral degree in 1963 (thesis in group theory). He was assistant professor for statistics at the University of Bonn between 1965 and 1972. His Habilitation 1969 was on stochastic models in demography. He published numerous papers on population mathematics, applied optimal control and the dynamics of the firm. He is author and editor of 7 books on demometrics and economic applications of control theory. Recently he wrote a comprehensive monographie on optimal control of economic processes (coauthored by R.F. Hartl).

VOLKER FIRCHAU (* 1949) is Privatdozent at the Department of Economics, University of Augsburg. He received the Diplom-Mathematiker degree from the University of Münster (1974), whereas he passed his doctoral promotion (1978) and his Habilitation (1984) in Augsburg. His current interest of research include decision theory, problems of insurance, and statistical methods. Two monographies on advanced topics on informational economics were recently published.

FRIEDRICH FRICKE (* 1928) is a member of the Vorstand of the Süddeutsche Treuhand Gesellschaft AG in Munich and is lecturing on Auditing and Taxes at the University of Ulm. He got his Diplom-Kaufmann degree from the University of Munich in 1951, and he is holding the additional degrees of Qualified Auditor (WP) and of Tax Advisor (StB). Friedrich Fricke published articles and comments on the German Accounting Legislation. His current fields of interest include auditing problems related to insurance companies and statistical methods applied to auditing.

HELMUT FUNKE (*1947) is Privatdozent at the Department of Economics, University of Ulm. He received his diploma degree 1973 from the University of Würzburg, whereas he passed his doctoral promotion 1977 and his Habilitation 1985 at the Faculty of Economics of the University of Karlsruhe. His papers cover topics such as price theory, economic indices, aggregation problems, and income-tax.

MARTIN GAYNOR (*1955) is Senior Economist at the Center for Health Economics Research in Needham, Massachusetts, U.S.A., and Faculty Research Fellow at the National Bureau of Economic Research, Cambridge, Massachusetts, U.S.A. He has taught economics at several universities in the United States, including the State University of New York at Binghamton, and Virginia Polytechnic Institute and State University. He received the B.A. and Ph.D. degrees from the University of California, San Diego and Northwestern University, respectively. His primary research interests are in the economics of organizations and the economics of health care.

ANDREAS GRUBER (*1960) studied mathematics and economics at the University of Ulm/FRG and the University of Wisconsin-Milwaukee/USA. He got his Diplom-Wirtschaftsmathematiker degree at the University of Ulm in 1985. Since that time he has a research and teaching position in the Department of Economics at the University of Ulm. His current field of interest cover finance, market equilibria, and information economics.

THOMAS HARTMANN-WENDELS (* 1957) is Wissenschaftlicher Mitarbeiter at the Lehrstuhl für Finanzierungslehre, Universität zu Köln. He received his Diplom-Kaufmann degree (1981) and his doctoral degree (1985) from the University of Cologne. In 1986 he was an exchange researcher at the School of Organization and Management, Yale University. His publications deal with advanced topics of capital market theory and his current interests of research are informational aspects in the theory of finance and in accounting.

MARTIN HELLWIG (* 1949) is currently Professor of Economics at the University of Bonn. He studied economics and history in Marburg and Heidelberg (Diplom-Volkswirt 1970) and at the Massachusetts Institute of Technology (Ph.D. 1973). He has held research and teaching positions at Stanford and Princeton Universities and at the University of Bonn, where he has been since 1977. His research interests concentrate on monetary theory and on the economics of information, in particular in application to financial and insurance markets.

THOMAS KAULMANN (* 1958) is research assistant at the Technical University of Munich, Institute for Business Administration and Industrial Engeneering. He studied economics at the University of Hannover, where he received his Diplom Oekonom degree in 1982. Occupations as research assistant at the Institute for Statistics and Econometrics and the Institute for Management and Organization at the University of Hannover followed. His 1986 doctoral dissertation was an empirical work on the property rights-theory of the firm. Current areas of interest for research include transaction cost economics, corporate innovation, entrepreneurship and office automation through new information technologies.

PAUL KLEINDORFER (*1940) is Professor of Decision Sciences and Economics at the University of Pennsylvania. Dr. Kleindorfer received his Ph.D. from Carnegie-Mellon University in 1970, and has taught at the Massachusetts Institute of Technology before coming to The Wharton School at the University of Pennsylvania. Dr. Kleindorfer has also been a visiting scholar at the International Institute of Management, West Berlin, at the International Institute for Applied Systems Analysis, Vienna, and, in the summer of 1986, at the University of Ulm. His primary research interests are incentives, regulation and managerial economics.

JAN PIETER KRAHNEN (*1954) is Hochschulassistent at the Faculty of Economics, Freie Universität Berlin.He studied in Paris (Université Dauphine, 1979), Rochester (Graduate School of Management, 1981/82) and Frankfurt (Johann-Wolfgang-Goethe Universität), where he graduated (Dipl. Kfm., 1980 and Dr.rer.pol., 1984). Since 1980 he teaches financial economics in Berlin.Current work concerns research on financial intermediation and the theory of investment, as well as consulting assignments on financial development.

GEORG MERAN (*1953) is wissenschaftlicher Mitarbeiter at the Faculty of Economics, Freie Universität Berlin, where he received his doctoral degree (1985). After studying political science and economics at the University of München he got the diploma degree (Dipl.Vw., 1978) at the University of Konstanz, where he was employed as wissenschaftlicher Mitarbeiter.Current work and interest are in the field of industrial and financial economics,enviromental economics and the theory of labor managed firms (comparative economics). He has authored (and co-authored) various papers in these fields.

HELLMUTH MILDE (*1941) is Associate Professor of Finance at the Faculty of Management, McGill University, Montreal (Quebec), Canada. He received his degrees from Hamburg University (Dipl. Volkswirt 1968) and Konstanz University (Dr.rer.soc. 1973; Habilitation 1979). He has been a visiting professor at the University of California, Los Angeles. He is author and co-author of several articles on commercial banking and credit markets in scholarly journals. A (German written) monograph on commercial banking behavior (co-authored with Ernst Baltensperger, Bern University) will be published shortly by Springer-Verlag.

ANDREAS PFINGSTEN (*1957) is working for the Institut für Wirtschaftstheorie und Operations Research (Universität Karlsruhe) and for the Landeskreditbank Baden-Württemberg. He studied in Karlsruhe (Diplom-Wirtschaftsingenieur 1982) and at the University of British Columbia, Vancouver B.C., Canada. In 1985 he got his doctoral degree in economics at Karlsruhe. His thesis on "The Measurement of Tax Progression" as well as a number of papers in the fields of index numbers, natural resource economics, and public finance were published in the past few years.

ERICH SCHANZE (* 1942) studied law at Frankfurt, Edinburgh and Harvard Universities. He received his first and second law degree in 1967 and 1972 respectively, a Master of Laws at Harvard University in 1969 (thesis on information liability), his Dr. iur. in 1975 (thesis on one-man corporation and piercing the corporate veil) and the Habilitation in 1984 at Frankfurt. He teaches civil law, business law, law and economics and comparative law in various capacities since 1972, including visitorships at Stanford Law School in 1978 and at the University of Chicago Law School in 1980. Since 1987 he is a Professor of Law at Ruprecht-Karls-Universität Heidelberg. His recent books and articles focus on international investment, long-term contracting, regulation, and law and economics.

REINHARD H. SCHMIDT (* 1946) studied business administration and economics in Frankfurt (Main). He received his degrees (Diplom-Kaufmann, 1971, Dr. rer. pol., 1974, and Habilitation, 1980) at the University of Frankfurt. In 1975/76, he was a visiting scholar at the Graduate School of Business, Stanford, Cal.. Since 1983 he is full professor of business administration and finance at the University of Trier (Fachbereich IV). His books and articles cover topics in financial economics, methodology and decision science.

DIETER SCHNEIDER (* 1935) is professor of business economics, especially taxation at the Ruhr-Universität Bochum. He got three degrees from the university of Frankfurt: Diplom-Kaufmann 1957, Dr. rer. pol. 196o, Habilitation in business administration 1965. Since 1965 he was full professor in Münster, since 197o in Frankfurt, and since 1973 in Bochum. His publications in professional journals as well as his books deal with various topics esp. of taxation, accounting, finance and investment, and the history of microeconomics and business economics.

GERHARD SORGER (* 1961) is research assistant at the Technical University of Vienna, Institute for Econometrics and Operations Research. He received both the Dipl.Ing. (1983) and the Dr.techn. (1986) degrees at the Technical University of Vienna. In his doctoral thesis he investigated the influence of a reference price on the optimal marketing strategies of a firm. His current interests of research are control theory and its applications to problems of economics.

KLAUS SPREMANN (* 1947) studied mathematics at the Technical University Munich, where he received his Diplom-Mathematiker degree in 1972 and where he wrote his doctoral dissertation on dynamic optimization in 1973. His Habilitation was at the Department of Economics at the University of Karlsruhe in 1975. Since 1978 he is full professor of Economics at the University of Ulm. Klaus Spremann was Visiting Professor at the University of British Columbia, Vancouver, in 1982 and at the National Taiwan University, Taipei, 1987. He wrote a book on modern finance, co-edited volumes on capital market theory, and published numerous papers. His main fields of interest are risk theory and information economics.

PETER SWOBODA (* 1937) is Professor at the University of Graz, Austria. He studied at the University of Commerce, Vienna, where he got the PH.D. in 1959. After the Habilitation (1964, University of Commerce, Vienna) he became full Professor at the University of Frankfurt, BRD. 1970 he moved from the University of Frankfurt to the University of Graz. He was Visiting Associate Professor at the University of Illinois, Urbana-Champaign (1965-1966). In 1976 he declined an offer from the University of Hamburg, BRD. He published two books on finance and numerous articles on cost accounting, capital budgeting, taxes and finance.

JOACHIM VOELLER (*1946) is professor of Economics at the University of Ulm. He received three degrees from the University of Karlsruhe: Diplom-Wirtschaftsingenieur (1971), Dr. rer. pol. (1974), Habilitation in Economics (1981). In 1982 he became professor of Economics at Karlsruhe and since 1985 he teaches in Ulm. His main fields of interest are: Theory of index numbers, money and credit, "Ordnungspolitik".

ALFRED WAGENHOFER (* 1959) is assistant professor of Business Administration at the University of Technology of Vienna. He received his Magister degree (1982) and his doctoral degree (1984) at the University of Vienna. His publications deal mainly with accounting and taxation issues, current interests of research are applications of agency models in accounting and auditing.

JOCHEN E.M. WILHELM (* 1945) is currently full professor of finance at the University of Passau. He studied mathematics and economics at the University of Bonn where he got his Diplom-Mathematiker degree in 1970; his doctoral dissertation 1973 was on multi-objective decision making under uncertainty, his Habilitation 1981 on dynamic capital market theory. 1982 - 1984 he taught as professor of business administration at the University of Bonn. He published three books and numerous papers on decision theory and on various topics in finance and banking.

Author Index

Subject Index

G. Bamberg, K. Spremann (Eds.)

Capital Market Equilibria

1986. 4 figures, 12 tables. X, 228 pages.
ISBN 3-540-16248-8

Contents: *G. Bamberg, K. Spremann:* Prologue. - *G. Bamberg:* The Hybrid Model and Related Approaches to Capital Market Equilibria. - *V. Firchau:* Portfolio Decisions and Capital Market Equilibria under Incomplete Information. - *R. Geske, S. Trautmann:* Option Valuation: Theory and Empirical Evidence. - *B. Rudolph:* The Value of Security Agreements. - *E. Schwartz, M. Brennan:* Asset Pricing in a Small Economy: A Test of the Omitted Assets Model. - *K. Spremann:* The Simple Analytics of Arbitrage. - About Contributors. - Author Index. - Subject Index.

This book contains introductory surveys and research results on the following six basic issues of modern business finance: decision-theoretic approaches to capital market equilibria, portfolio decisions under incomplete information, option valuation (theory and empirical evidence), the value of security agreements, asset pricing and testing, and arbitrage analysis.
The aim of this book is to bridge the gap between textbooks and professional journals in the field of business finance. The volume covers modern concepts such as heterogeneous and incomplete information, arbitrage theory, and contracting on security agreements, which are rarely found in textbooks. Furthermore, the topics of the book were selected in order to give a sound and diversified presentation of current lines of knowledge in the theory of business finance. The book reflects the cooperation between eight authors from the USA, Canada, and West Germany.

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo

Springer